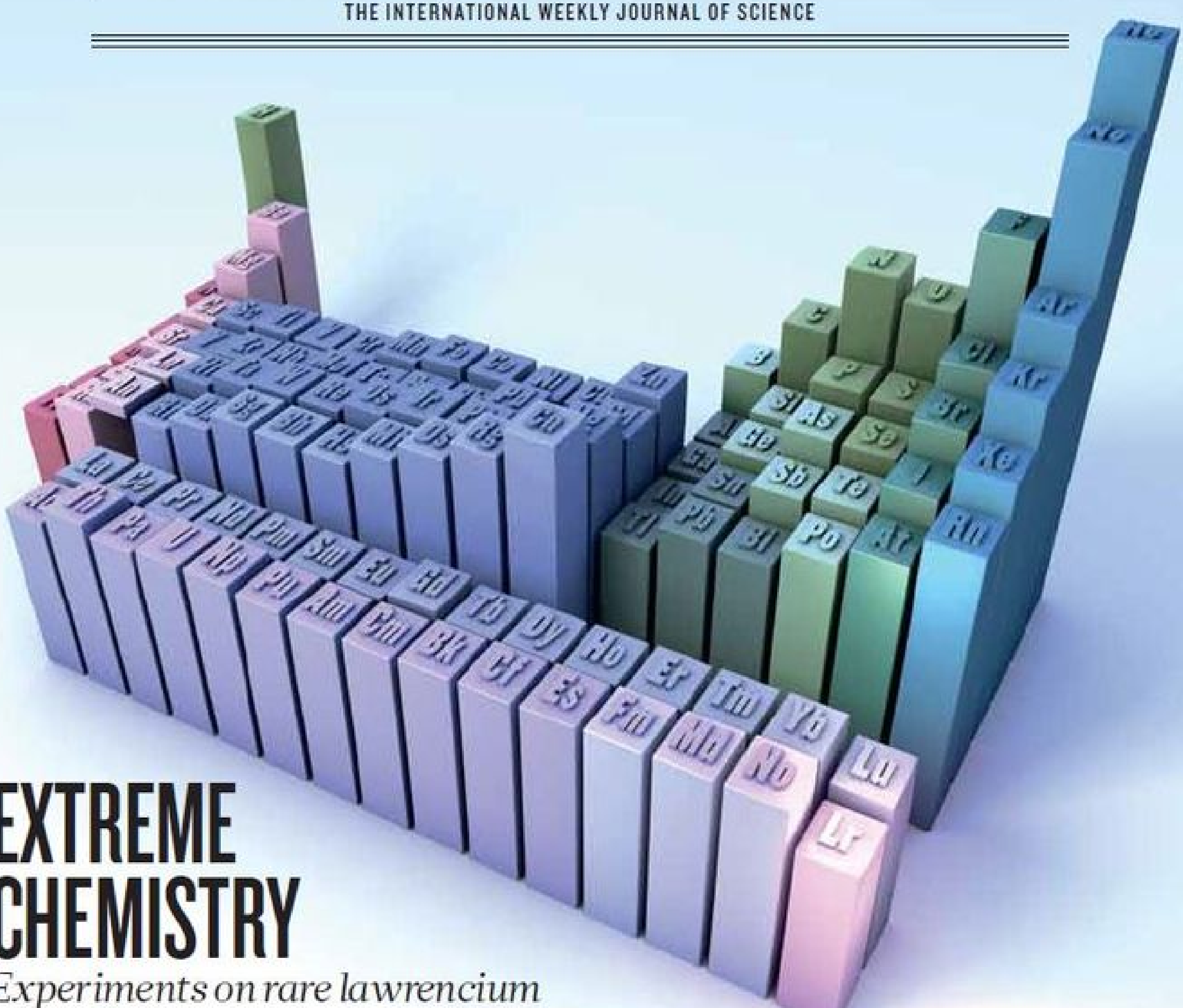
nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



EXTREME CHEMISTRY

Experiments on rare lawrencium atoms illuminate a relativistic region of the periodic table **PAGES 166 & 209**

RESEARCH CAREERS

POSTDOC OF THE FUTURE

Why permadocs must become superdocs

PAGE 144

BIOINFORMATICS

BIOLOGY BY THE NUMBERS

Data analyis is undervalued in the life sciences

PAGE 151

ARCTIC PERMAFROST

COLD COMFORT

Green-house gas release may be slow but prolonged

PAGE 171

NATURE.COM/NATURE
9 April 2015 £10
Vol. 520, No. 7546

THIS WEEK

EDITORIALS

MOON SHOT Fresh estimates support cosmic-collision origin **p.132**

WORLD VIEW Test aircraft engines to benefit from volcano science **p.133**

APRON STRINGS Hundred-million-year-old insect mothered young **p.134**

All that glitters

A review of the United Kingdom's progress towards 'gold' open-access research is instructive — for funders, publishers and scientists both at home and abroad.

In 2012, the United Kingdom took a bold leap on open-access publishing, announcing that all research articles produced by its publicly funded scientists should be made free to read. A fine pledge, but three years on, it has experienced some practical difficulties. It is instructive to examine them.

To quote the mantra of Research Councils UK (RCUK), the umbrella body for the seven national funding agencies that is overseeing the publishing conversion: "open access is a journey, not an event." Continuing that metaphor, it seems that the United Kingdom has been sent out as an advance party on this journey. Its scientists and publishers are scouting through thickets of confusion on their way to bringing about 'gold' open access. This is the system in which a published article is immediately made freely available, with maximum opportunity to reuse it for applications such as text-mining and translation.

Many nations have not set open-access policies. Others, including the United States, are loitering with little intent, and mandating only delayed access to an author's version of a peer-reviewed manuscript — a 'green' form of open access that ultimately benefits science less (see *Nature* **494,** 401; 2013). RCUK favours a mixed model, but one that gradually migrates towards gold. A review of its progress, published in March, serves as a useful guide and should be examined by funders, publishers and institutions (see go.nature.com/tz2orl).

One problem is that it is hard to track progress, good or bad. RCUK and many British institutions cannot systematically count RCUK-funded papers, let alone those published as open access. As a result, RCUK, although strongly confident, cannot be entirely sure whether the £17-million (US$25-million) open-access fund it gave to universities in 2013–14 has produced the desired result of at least 45% of its funded papers being either green or gold open access.

This underlines the need for researchers to use the ORCID system, a single digital identifier for individuals that links their published papers and grant applications. Use of FundRef, a service from non-profit publisher alliance CrossRef for reporting funding sources, is also essential.

Open-access licences are another major source of confusion. The London-based biomedical charity the Wellcome Trust, which has long mandated gold open access and provides the funds to cover it, reported last month that it now sees 87% compliance with its policy — but that only 66% of papers are accompanied by a liberal publishing licence that allows extensive reuse of text. Licence information, it says, is often ambiguous or contradictory, and records for open-access payments can be lost between authors and publishers.

RCUK says that the licence problem is compounded by researchers not understanding which licence they need to use to comply with the open-access policy, and by publishers offering a range of 'open' licences. (Since January, all 18 open-access journals owned by Nature Publishing Group have switched to using the fully liberal CC-BY 4.0 licence as a default, and to charging a flat fee.)

And then there are costs. All experiments should be encouraged in the evolving gold open-access market, but academics should know that fees for papers published in fully open-access journals are lower than those of 'hybrid' subscription journals that allow an open-access option. The Wellcome Trust says that the average fee levied by hybrid journals is 64% higher than that charged by fully open-access titles. British funders are now pondering steering the market by dissuading researchers from publishing in hybrid journals, as other countries have done.

"Britain looks increasingly isolated in its gold-leaning stance."

The RCUK review did not have the remit to question whether RCUK should continue to hand out money for gold open-access publishing. But with a new UK government in the offing and the country looking increasingly isolated in its gold-leaning stance, there must be a concern that the agency might end up scrapping its gold preference. Last year, four influential UK university-funding bodies announced a green open-access policy that will further steer academics towards delayed public archiving of manuscripts.

In conclusion: the road to gold open access will be bumpy and hard to navigate. But what is encouraging is that these issues are being aired, with publishers, funders and researchers talking to each other about the costs and challenges. The take-home message from the RCUK review is the need to keep discussing difficulties publicly — for only then can other nations learn from the United Kingdom's experience. ■

Seeds of change

The European Union faces a fresh battle over next-generation plant-breeding techniques.

The US plant-breeding company Cibus is proudly rolling out its first crop created with an innovative precision gene-editing technology: herbicide-tolerant oilseed rape.

The crop will be planted in the United States this spring and the firm already has authorization to cultivate it in Canada. The technology switches just a few nucleotides in a plant's DNA; the company's webpage points out that because it works without integrating foreign genetic material, the resulting plants cannot be stigmatized as transgenic. They will, it optimistically declares, "be globally acceptable".

Cibus, based in San Diego, California, hopes that plants imbued in this way with traits that improve their robustness or nutritional value will also find favour in the European Union (EU), where many countries vehemently oppose genetically modified (GM) crops created by transfer of specific foreign genes.

That hope has logic on its side, and it is not misplaced. In February, authorities in GM-hostile Germany told Cibus that they would not consider products created by gene editing as GM, but as products of conventional plant breeding. However, with new battle lines already being drawn, broader approval and acceptance are unlikely to be so simple.

The first battle for GM crops in Europe is currently drawing to an unsatisfactory close. EU legislation from 2001 dictates that the European Food Safety Authority (EFSA) must carry out a scientific risk assessment of any GM strain submitted for authorization. Member states must then vote on whether to pass it, obliging all of them to permit the crop's cultivation if it is approved.

But these votes have almost never yielded the required majority for or against any strain given the EFSA green light. And the European Commission has never dared to exercise its right to enforce a positive decision in cases of impasse. Instead, it has proposed new rules, which came into force last week, allowing EU member states to opt out of a requirement to allow cultivation on non-scientific grounds.

Although it is smart politics, the rule will not be enough to break through the authorization impasse, because all nations must still vote, and a qualified majority must be reached. So, in the next few weeks, the commission will propose further legislation that is likely to allow member states to opt out of the authorization process too. This could finally get the system going again, and give GM-friendly countries such as Spain a wider range of GM crops to choose from.

Meanwhile, science has moved on. Plants without foreign genes can now be created with a variety of methods and technologies that precisely tweak or change the regulation of a native gene. Such plants should reassure anti-GM lobbies that criticize the moral right of scientists to 'play God', and the alleged instability of foreign genes. But environmental groups such as Greenpeace seem far from convinced. In January, several groups wrote an open letter to the commission insisting that new methods that change DNA structure or interfere with gene regulation in any way should also be subject to the EU's tight GM regulations. They argue that the precautionary principle should continue to apply — and that because of the enhanced abilities of the technologies, the safety bar should in fact be raised.

The commission is again playing for time. In 2007, it appointed an expert panel to advise it on the ever-expanding plant-breeding toolbox. The panel's report was submitted in 2012 but never published. The commission now says that it has set in motion a "thorough legal analysis" of the definition of 'GM organisms' in its own legislation, and of the criteria for excluding certain technologies. The conclusions of the analysis, it warns, "cannot be anticipated".

"The European Commission is again playing for time."

Most plant scientists consider the new tools to be helpful extensions to normal plant-breeding practice. In many cases, they say, the plant products are indistinguishable from the original plants and are intrinsically even safer than GM plants. Two years ago, the European Academies Science Advisory Council, an umbrella group for national academies in Europe, argued that the time had come for regulators to abandon their fixation on plant technologies and instead carry out risk assessments on the individual plant products. In February, the European Plant Science Organisation, an independent body representing more than 220 research institutes and universities from 28 European countries, as well as Australia, Japan and New Zealand, reiterated the message.

Late last month, the Leopoldina, Germany's national academy of sciences, published a similar position paper in the hope of influencing its government, as the country deliberates anew the legal environment of its GM regulations. Both Germany and the commission are watching and waiting. In their letter to Cibus, the German authorities noted that their statement of readiness to consider products of gene editing as non-GM would be invalidated if the European Commission were finally to decide otherwise.

As Europe's first battle on GM staggers to an uneasy truce, a second — and perhaps more important one — is approaching. ■

Lunar affairs

A study in Nature *adds a dramatic twist to the backstory of a neighbour we thought we knew.*

In the stories of many human cultures, the Solar System is something of a family affair. The Norse people and the Incas of South America believed that the Sun and the Moon were brother and sister. A Native American myth has them as husband and wife (although the husband wants to eat his children, the stars). The Moon as a mother is a common theme.

Now scientists have suggested a rival celestial tale with a twist that is more common to terrestrial television dramas: the sudden appearance of a long-lost sibling. The early Earth had a near-identical geological twin, they say. The two young planets, of course, had a falling out and the twin vanished. But before it did so, it saddled Earth with an orbiting Moon-child.

The the origin of the Moon is a classic story that has been told many times. The latest version — written in a research paper on page 212 — still has some plot holes. But it is a cracking tale.

The time: some 4.5 billion years ago, in the earliest days of the Solar System. The place: hostile. A long time ago, if you like, in a galaxy not very far away. Thousands of adolescent protoplanets whizz around the Sun, bashing into each other, some breaking into smaller pieces and forming others as they soak up the freed materials. One of these protoplanets, lying not too far from the Sun and not too close, is what we now call Earth.

Enter, stage left, protoplanet Theia. Smaller than proto-Earth, it was raised in a similar neighbourhood. A chance meeting set the two on a collision course. The meeting is violent, and — here it helps if you imagine the most gravelly cinematic voice-over you can manage — life for both will never be the same again.

Theia becomes a giant cloud of dust infused with bits spewed from the injured proto-Earth, which quickly comes together to form the Moon. Earth gains a dependant.

The script might sound familiar; the plot more of a remake than of anything original. But here is the difference. Previously, many planetary scientists considered that it was too much of a stretch to say that young Theia and Earth were so closely related. It is much more plausible, given the chaos of the time, to present Theia as a random stranger. But that creates a continuity error: the mineral composition of rocks retrieved from the Moon is eerily similar to those of Earth.

Either Theia and Earth are related, or our best models of how the Moon formed are wrong. But if they are related, then why is it that the other bodies in the Solar System that we have studied seem to be so different from each other? What are the chances, given the number of objects out there at the time, that proto-Earth would be hit by a near clone?

The latest study runs computer simulations of those early days, to investigate the possible backstories of the major characters, including how and where they formed and their probable orbits. (A note to film directors: this bit is probably best presented as a montage.) The number crunching offers a realistic script: there is a one-in-five chance that proto-Earth and Theia could have formed at about the same distance from the Sun, so from the same stuff, and then run into each other to make the Moon. True, it is not a cut-and-dried ending that ties up all the loose ends before the credits roll. But all the best stories leave room for a sequel. ■

➲ NATURE.COM
To comment online, click on Editorials at: go.nature.com/xhunqv

Test the effects of ash on jet engines

To judge the safety of flying during an eruption, the airline industry cannot just rely on advances in volcanic monitoring and prediction, says **Matthew Watson**.

Next week marks five years since the eruption of the Icelandic volcano Eyjafjallajökull that halted air traffic over Europe. For six days flights were grounded as politicians and the public wrestled with ideas of risk management and scientific uncertainty. Good decisions were made at the time, but in the immediate aftermath, governments and airlines promised that they would learn the lessons of a shutdown that cost the global economy an estimated US$5 billion. So would things be different if the same eruption happened tomorrow? Yes and no.

We were unlucky in 2010. The magma produced by the volcano was rich in silica, making the lava very viscous, and it emerged amid abundant glacial melt water. As a result, the ash particles were unusually fine, and travelled far into the heart of Europe on (very inconvenient) stable northerly winds.

There have been two eruptions in Iceland since 2010: Grímsvötn in 2011 and Bárðarbunga in 2014–15. Neither had the same impact, partly because the meteorological and geological conditions were more benign. But partly also because our ability to monitor ash clouds, and to predict how they disperse, has improved. These improvements have fed into better management of air space, and ultimately helped us to keep more planes flying.

> DECISION-MAKERS ARE BETTER **INFORMED** OF THE HAZARD AND POSSIBLE **RESPONSES** THAN THEY WERE IN 2010.

The biggest policy change since 2010 is a shift from a decision-making process that responded simply to the presence of any ash, to one based on a safe threshold. Aircraft can now fly as long as the ash concentration does not exceed 0.2 milligrams per cubic metre.

This safety limit demands more accurate satellite measurements of atmospheric ash than were available in 2010. Frustratingly for those of us in the field, the knowledge existed to make these measurements five years ago, but was not used. In fact, the basic satellite algorithm to detect ash levels in 2010 was already 20 years old. It takes time to translate research findings into applications, but it is a salutary lesson that it took a crisis to provoke the political effort to upgrade to using more-recent research.

Other rapid advances have been made since 2010 which, coupled with the (better late than never) uptake of existing algorithms, have led to a step change in our ability to remotely monitor eruptions and to track the concentration and height of ash.

Dispersion models have improved. One example is PlumeRise, a fluid dynamical model that corrects for the interaction of wind with the eruption column. This helps researchers to calculate more accurately the intensity of the eruption, which improves model predictions. Models used during the 2010 crisis did not account for the bending of the column in high wind, and potentially underestimated the amount of material injected into the atmosphere. The online version of this model is now used by volcanic ash advisory centres around the world — directly as a result of the Eyjafjallajökull event.

➲ NATURE.COM
Discuss this article online at:
go.nature.com/9tsdve

There has also been major investment in hardware. An array of state-of-the-art ground-based sensors has been deployed in Iceland to measure both volcanic ash and gases. Here in the United Kingdom, the government has invested in a Lidar network to monitor ash from the ground, and a second aircraft specifically to check levels from the skies.

Decision-makers are better informed of the hazard and possible responses than they were in 2010. For example, the risk that volcanic ash poses to airspace and infrastructure is now captured in the national risk register, which lays down scenario plans for different types and scales of volcanic activity. And the eruptions of Grímsvötn and Bárðarbunga provided an opportunity to test and improve these responses.

The Civil Aviation Authority (CAA) has developed new rules for flying in ash. UK airspace is now divided into three areas of density: low, medium and high. All aircraft may fly in low-density areas but airlines are required to set out how they intend to fly their fleet safely through medium- and high-density areas and have these safety cases approved by the CAA. The CAA states that many safety cases are already in place for medium-density areas.

Does that all mean that airspace will never again be shut, as it was in 2010? Probably, but with a couple of important caveats. First, some events will have an impact no matter how well managed. A very large, ash-bearing eruption would still halt air traffic.

Second, despite being able to monitor, measure and predict the concentrations of ash more adroitly, considerable uncertainties remain. The largest of these is how tolerant jet engines are to ash, particularly the newer and hotter engines.

Very few tests have been conducted to see how engines cope with ash. Such experiments are expensive and complex to run, especially on whole engines (rather than just on components). Yet, without a clearer idea of that tolerance, sharp delineations of the acceptable areas to fly in, such as those made by the CAA, are unrealistic. Some experiments are planned. Until then, airlines are likely to struggle to make a convincing safety case to fly in medium or high densities of ash.

Without that effort from the industry, a lot of the scientific progress made will be hard to use during the next ash-cloud crisis. ■

Matthew Watson *is reader in the School of Earth Sciences and Fellow of the Cabot Institute at the University of Bristol, UK.*
e-mail: matt.watson@bristol.ac.uk

MARINE BIOLOGY

Sea creatures adapt to acid

Sea urchins can radically alter their energy use to cope with more-acidic oceans.

Donal Manahan led a team at the University of Southern California in Los Angeles that grew *Strongylocentrotus purpuratus* urchins in current seawater conditions and in more-acidic conditions that are expected under some climate-change scenarios. They found no difference between the two larva groups in terms of size, gene expression or metabolic rate. But larvae feeding in the more-acidic water allocated 84% of their ATP, which transports energy within cells, to protein synthesis and ion transport, whereas larvae feeding in normal conditions allocated just 55% of their ATP to these tasks.

Altering their metabolism could help sea urchins and other marine organisms to withstand climate change, the authors say.

Proc. Natl Acad. Sci. USA **http://doi.org/3cg (2015)**

JIM BRANDENBURG/MINDEN PICTURES/FLPA

GLACIOLOGY

Few Canadian glaciers left by 2100

Mountain glaciers in western Canada could shrink by 70% relative to 2005 levels by the end of the century as a result of global warming.

Garry Clarke of the University of British Columbia in Vancouver, Canada, and his colleagues built a high-resolution model that includes ice dynamics and then ran it with a series of climate scenarios covering the twenty-first century. The model suggests that few glaciers in the Canadian Rocky Mountains (pictured) will persist by 2100, although glaciers in the coastal range of north-west British Columbia could survive "in a diminished state", say the authors.

The team predicts that changes in run-off from the melting glaciers over the course of the century could affect aquatic ecosystems, agriculture, forests, water quality and tourism.

Nature Geosci. **http://dx.doi.org/10.1038/ngeo2407 (2015)**

PALAEONTOLOGY

Mesozoic insect mothering

Insects that care for their young have been around for at least 95 million years, the discovery of an amber-entombed bug from Myanmar suggests.

B. WANG *ET AL.*

All kinds of extant insects care for their offspring, including some species of scale insect (Coccoidea) that hatch their young from egg sacs on their abdomens. Bo Wang at the Nanjing Institute of Geology and Palaeontology in China and his team now provide evidence for such behaviour in a new species of scale insect, *Wathondara kotejai* (**pictured**), which was found in northern Myanmar in a piece of amber 95 million to 105 million years old. Approximately 60 eggs coat the abdomen of the entombed female specimen, and six newly hatched individuals were also preserved in the amber.

eLife 4, **e05447 (2015)**

AVIAN BIOLOGY

Small bird takes big journey

An extraordinary feat of migration has been confirmed in a tiny songbird that weighs just 12 grams.

Blackpoll warblers (*Setophaga striata*) have long been thought to fly non-stop from northeastern North America to the Caribbean or South America. William DeLuca of the University of Massachusetts, Amherst, and his colleagues fitted the birds with 0.5-gram devices that record light levels over time, allowing latitude and longitude to be inferred from the date and the timing of dusk and dawn. Five birds were successfully recaptured, and the data suggested that the birds flew in a straight line over the Atlantic Ocean to the Caribbean, where they made a stop before continuing to wintering grounds in South America.

The journey required three days of non-stop flight and covered some 2,500 kilometres — one of the longest-recorded migrations for a bird of its size, the authors say.

Biol. Lett. **http://doi.org/3ch (2015)**

NEUROSCIENCE

How to form a fake memory in mice

Unrelated memories can be artificially linked by activating distinct groups of neurons at the same time.

Kaoru Inokuchi at the University of Toyama, Japan, and his colleagues let mice explore a cylindrical container. Later they placed the mice in a cubic box and gave them an immediate electric shock to the foot. The scientists then used light to stimulate those neurons in the hippocampus and amygdala regions of the mouse brain that had been activated both when the animals had laid down their memories of the safe cylindrical container and when they received the electric shock in the box.

When light-stimulated mice were placed back in the cylinder, they froze for a longer time than mice that did not receive brain stimulation, suggesting that the stimulated mice connected the shock with the safe environment.

Cell Rep. **http://dx.doi.org/10.1016/j.celrep.2015.03.017 (2015)**

ROBOTICS

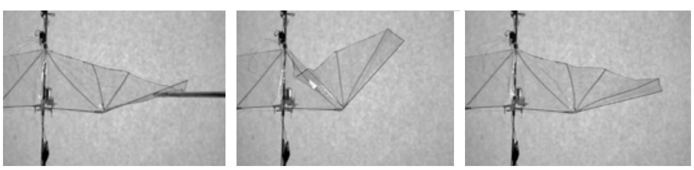
Wings that cope with collisions

A wing inspired by those of birds and bats can overcome collisions with obstacles, and might allow the development of flying robots that have improved energy efficiency.

Amanda Stowers and David Lentink of Stanford University, California, built a robotic flapping wing consisting of an 'arm' attached to the body and a hinged 'hand' attached to that. The design naturally unfolds when the wing beats during flight. Modelling revealed that flapping generates acceleration, unfolding the wing.

The researchers showed in the lab that when the wing's hand hits a branch (**pictured**, left), it compensates by furling in (middle) and then unfurling (right), surviving impacts of up to 5 metres per second. This design could help future flapping robots to deal with tricky flight paths without the need for complicated algorithms, say the authors, because collisions could be survived rather than avoided.

Bioinspir. Biomim. 10, **025001 (2015)**



LENTINK LAB/IOP PUBLISHING

GENOME EDITING

CRISPR controls gene expression

A genome-editing technique can be modified to switch specific genes on and off through 'epigenetic' changes.

The technique, called CRISPR, allows biologists to edit selected DNA sequences, but it does not normally alter the epigenome — the chemical modifications to DNA and associated proteins that regulate gene expression. To target the epigenome, Timothy Reddy and Charles Gersbach at Duke University in Durham, North Carolina, and their colleagues mutated the enzyme used for CRISPR so that it no longer cut DNA, and fused it to part of an enzyme that transfers acetyl groups onto proteins associated with DNA.

The researchers targeted the modified enzyme to proteins near specific DNA sequences, and showed that the added acetyl groups boosted the expression of associated genes. The approach could be used to understand how specific epigenetic changes affect gene expression near the site of chemical modification and farther away, the authors say.

Nature Biotechnol. **http://dx.doi.org/10.1038/nbt.3199 (2015)**

SOCIAL SELECTION

Popular articles on social media

Papers need summaries for the public

In a bid to make research papers more accessible to non-scientists, journals should require scientists to write lay summaries of their articles. That is the proposal put forward in a recent opinion piece that caught the attention of researchers on social media. Such summaries would clearly explain the importance and implications of the work, and drew support from some commenters online. "Lay summaries needed to enhance communication, visibility and impacts of scientific research," tweeted Jiangxiao Qiu, an ecology PhD student at the University of Wisconsin-Madison. But Björn Brembs, a neuroscientist at the University of Regensburg in Germany, tweeted: "Such summaries sure are beneficial, but 'needed'? Written by the authors? Not so sure."

Proc. Natl Acad. Sci. USA 112, **3585–3586 (2015)**

Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

NATURE.COM
For more on popular papers:
go.nature.com/ti9wqn

ARCTIC SCIENCE

Bacteria warm up permafrost

Heat produced by microbes that break down organic material can speed up the thawing of frozen soils in the Arctic.

Thawing permafrost in a warming Arctic could release large amounts of organic carbon into the atmosphere, enhancing global warming. Bo Elberling at the University of Copenhagen and his colleagues determined the amount of heat produced by microbial activity in permafrost soils sampled from six sites in Greenland. They found that different types of organic soil are all susceptible to ground heat production, whereas mineral soils are less prone to it.

Using a model of heat and water flow, the scientists estimate that, as increasing temperatures begin to melt the permafrost, microbial heat production will also increase, accelerating the thaw.

Nature Clim. Change **http://dx.doi.org/10.1038/nclimate2590 (2015)**

OCEAN SCIENCE

Ocean bacteria key to carbon sink

Bacteria in the ocean might be the main driver of processes that trap carbon in complex molecules in sea water.

Ronald Benner at the University of South Carolina in Columbia and his team placed bacteria from coastal sea water in a simulated ocean environment containing simple sugars. After 29 days, they observed that these sugars had been converted by the bacteria to complex molecules like those found in the ocean. The team suggests that bacteria are responsible for much of the carbon in such organic particles in the ocean, and that models of how atmospheric carbon is trapped should consider how rapidly bacteria can produce these compounds.

Nature Commun. 6, **6711 (2015)**

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

POLICY

Iran nuclear deal

On 2 April, Iran reached a preliminary agreement with the United States and other nations aiming to cut off its means of producing nuclear weapons. Among other changes, Iran would be required to slash the number of centrifuges it maintains for enriching uranium and to neutralize most of its stores of enriched uranium in exchange for relief from long-standing economic sanctions against the country. Negotiators aim to finalize the agreement over the next three months.

Academic freedom

Activists in Hong Kong are speaking out about perceived interference by the Chinese government in local academic institutions. More than 500 people signed a petition that ran on 30 March in the Chinese-language newspaper *Ming Pao*. Signatories expressed growing concerns about political influence on the appointments of university administrators and on choices of teaching materials. Activists have accused the Chinese government of cracking down on educational institutions following student-led pro-democracy protests in Hong Kong that lasted for 79 days in 2014.

MAX WHITTAKER/GETTY

Water cutbacks for California

As California's drought drags on, Governor Jerry Brown has ordered the state's first ever mandatory water restrictions. The plan, announced on 1 April, calls for measures to cut the state's non-agricultural water usage by one-quarter compared with 2013. Many of California's ski resorts have closed early this year because of low snow levels (pictured). On 1 April, for the first time in 75 years, surveyors had no snow to measure at an annual assessment at Phillips Station, a 2,073-metre-high site near Echo Summit, California. The water department measured only 3.6 centimetres of water content in the snowpack statewide — 5% of the historical average for 1 April. The snowpack provides about 30% of the state's fresh water. See go.nature.com/gcgumk for more.

Whales win in court

A US court ruled on 31 March that marine-wildlife officials violated environmental law by authorizing US Navy exercises that could harm marine mammals off the coasts of southern California and Hawaii. The decision states that the National Marine Fisheries Service — the branch of the National Oceanic and Atmospheric Administration (NOAA) that approved the activities — was wrong in concluding that there would be a negligible impact on whales and other animals. NOAA said in a statement that it was reviewing the judge's decision and evaluating possible next steps.

Plant patent flap

Agricultural-industry bodies have condemned a decision by the European Patent Office (EPO) to potentially allow patents on plant products such as fruits and seeds even if the plants had been created by conventional, non-patentable breeding techniques. The European Seed Association in Brussels said on 31 March that the move could restrict access to many plant products and might discourage innovation in plant breeding. The decision, made on 25 March by the EPO's Enlarged Board of Appeal, cannot be appealed.

Climate clampdown

On 31 March, the United States formalized a pledge to reduce its greenhouse-gas emissions to 26–28% below 2005 levels by 2025. Russia has committed to cut greenhouse gases to 25–30% below 1990 levels by 2030. Several other nations and the European Union filed their climate plans earlier this year to prepare for negotiations at the UN global-warming summit in Paris this December. See go.nature.com/L1iixk for more.

FUNDING

Megascope money

Canada will contribute about Can$240 million (US$193 million) towards construction of the Thirty Meter Telescope over the next decade, Prime Minister Stephen Harper said on 6 April. The announcement ends months of speculation about Canada's future role in the megatelescope, to be built on Mauna Kea in Hawaii at an expected cost of more than US$1 billion. Roughly Can$150 million of Canada's money will go towards building the enclosure to house the giant telescope. Project officials had expected

Canada to announce its commitment last year (see *Nature* **519,** 270–271; 2015).

EVENTS

LHC restarts

The Large Hadron Collider (LHC) kicked off its second run on 5 April, after a two-year hiatus for an upgrade. The world's most powerful particle collider will now run at a higher energy and produce more proton collisions. A short circuit late last month in a safety system had delayed the start-up, but it was resolved on 30 March by engineers at CERN, Europe's particle-physics laboratory near Geneva, Switzerland. See go.nature.com/gexdrq for more.

STD lawsuit

Johns Hopkins University, the Rockefeller Foundation and the pharmaceutical company Bristol-Myers Squibb were sued on 1 April for their alleged contributions to unethical experiments that exposed Guatemalans to sexually transmitted diseases (STDs) in the 1940s and 1950s (see *Nature* **482,** 148–152; 2012). The suit, which seeks more than US$1 billion in damages, was brought by 774 plaintiffs. A class-action lawsuit was filed against US officials in 2011 but was dismissed on the grounds of governmental immunity.

JEANNE HYDE/WHALE OF A PURPOSE

Killer-whale baby

Whale watchers caught the first glimpse of a few-day-old calf in a pod of endangered killer whales (*Orcinus orca*) near Active Pass, British Columbia on 30 March. Baby J52 (**pictured**) is the fourth whale born to the local population in the past four months, say scientists at the Center for Whale Research in Friday Harbor, Washington, which verified the sighting. The centre helps the US government to monitor 'southern resident' killer whales in the North Pacific Ocean. The population consists of three pods and an estimated 81 individuals.

Global positioning

China has taken its first step towards expanding its regional satellite-navigation system into one with global coverage, similar to the US Global Positioning System and Russia's GLONASS. On 30 March, China launched a new-generation satellite from the Xichang Satellite Launch Center in the Sichuan province. The satellite, launched on a Long-March-3C carrier rocket, forms part of the BeiDou Navigation Satellite System.

BUSINESS

Gene kits hit UK

DNA-testing kits, banned in the United States, can now be bought in UK stores. On 30 March, personal-genetics firm 23andMe in Mountain View, California, announced that its product is available in more than 600 locations of the UK shop Superdrug. 23andMe was ordered to stop marketing the service in the United States in November 2013, when the US Food and Drug Administration determined that there was insufficient information about its safety and effectiveness (see *Nature* http://doi.org/qdm; 2013).

No to deforestation

Global food-supplier Archer Daniels Midland confirmed on 30 March that it will stop buying soya beans and palm oil from producers involved in deforestation. The company, headquartered in Chicago, Illinois, will work with outside organizations to monitor compliance at farms from which it gets these products. The announcement follows similar pledges by major suppliers such as ConAgra Foods in Omaha, Nebraska.

COMING UP

12–17 APRIL
Vienna hosts the general assembly of the European Geosciences Union, which look at topics such as the environmental impacts of geoengineering, the results from the Rosetta probe and changes to the Greenland ice sheet.
go.nature.com/9manc8

13–16 APRIL
Topics such as ride-sharing technologies, small-satellite design and solar electric propulsion of spacecraft will be discussed at the 31st Space Symposium in Colorado Springs.
go.nature.com/tsn9ex

15–19 APRIL
Prehistoric dogs and three-dimensional modelling and printing are among the items on the agenda at the 80th annual meeting of the Society for American Archaeology in San Francisco, California.
go.nature.com/guxdpf

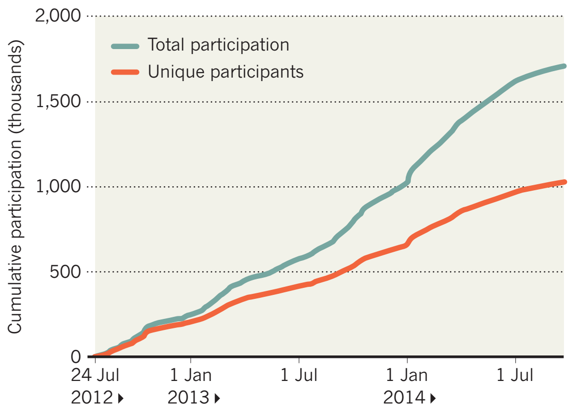
TREND WATCH

In one of the largest studies of massive open online courses (MOOCs), researchers at Harvard University and the Massachusetts Institute of Technology (MIT) in Cambridge report the steadily rising popularity of the non-profit learning platform launched jointly by the universities in 2012 (see go.nature.com/dk2ntt). Between July 2012 and September 2014, the authors clocked 1,300 people starting MOOCs every day, with each participant taking 1.7 courses, on average.

SOURCE: MIT/HARVARD

POPULAR PROGRAMMES

Open online courses are taking off at Harvard and the Massachusetts Institute of Technology.



NATURE.COM
For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

CLIMATOLOGY Climate scientists attempt to lure talented physicists **p.140**

BIOTECHNOLOGY Synthetic biologists write the rules for living machines **p.141**

AFRICAN SCIENCE Regional hub will provide research autonomy **p.142**

CAREERS What the future holds for a growing postdoc population **p.144**

AMI VITALE/FAO

Goats and sheep are sold frequently, which could challenge the vaccination effort against peste des petits ruminants.

DISEASE ERADICATION

Race to stamp out animal plague begins

Killer disease that is scourge of world's poorest ruminant farmers is ripe for elimination.

BY EWEN CALLAWAY

Humanity wiped out smallpox in 1980 and the cattle virus rinderpest in 2011. Polio stands on the brink of eradication, with just 21 cases recorded this year worldwide. Now, health officials have launched a global effort to vanquish yet another disease — a sheep- and goat-killer that is little known in rich countries, but creates economic ruin for the world's poorest people.

A conference hosted by the United Nations in Abidjan, Côte d'Ivoire, from 31 March to 2 April, marked the roll out of a global campaign to eradicate by 2030 the sheep and goat virus, which is known as PPR — an abbreviation of its French name, peste des petits ruminants. PPR eradication is technically feasible, say animal-health specialists, but it is uncertain where the effort's organization and billions of dollars of necessary funding will come from.

"This is an exercise in persuading the world community and funders that this work could and should be done," says Jeffrey Mariner, an epidemiologist at Tufts University veterinary school in North Grafton, Massachusetts, who attended the meeting.

PPR is related to measles and rinderpest, which once threatened the livelihoods of cattle herders, especially those in Africa. Causing high fever, diarrhoea and lesions in the mouths of sheep and goats, PPR is highly infectious and kills 30–70% of the animals it infects. It is endemic across northern, central and west Africa and south Asia, and it has more recently taken hold in China and Turkey.

The UN puts the economic costs of PPR at between US$1.5 billion and $2.1 billion per year, a burden borne by some of the world's poorest people, who rely on sheep and goats for food and income (see 'Cost of a goat-killer'). "Sheep and goats are the cattle of ▶

the poor, and they are the bank for the poor," says Bernard Vallat, director-general of World Organisation for Animal Health (OIE) in Paris, which co-hosted the meeting.

"Somebody in Kenya said to me, 'if the goats die, the children don't go to school'," says Michael Baron, a virologist at the Pirbright Institute in Woking, UK. Sheep and goat herding also helps many women in the developing world to attain self-sufficiency.

PPR ticks many of the boxes needed for an eradication campaign: an effective vaccine has been available for decades, and scientists have created formulations that remain effective for weeks without refrigeration. Diagnostic tools, including some that can be used in animal pens, are accurate. There also seems to be no wild reservoir from which the virus may rebound once wiped out from domestic flocks and herds. "It's one of those no-brainers," says Christopher Oura, a veterinary virologist at the University of the West Indies in St Augustine, Trinidad, who studies PPR. "The tools are out there to eradicate the virus."

A cost–benefit analysis by Mariner and his colleagues estimated that eradication will save more than $42 billion over a hundred years. The fresh success of the rinderpest eradication campaign also gives many hope that a second animal disease can be eliminated for good. "Before everybody forgets that, people want to get the rest of the world on board," says Baron.

PPR eradication presents its own challenges. The campaign strategy focuses on dramatically ramping up and coordinating vaccinations, but this is complicated because sheep and goats are more abundant than cattle in most of the developing world, and people hold onto them for a shorter time before selling or slaughtering them. The campaign will also attempt to systematically target areas where the virus is spreading, but veterinary services are weak in many of those regions.

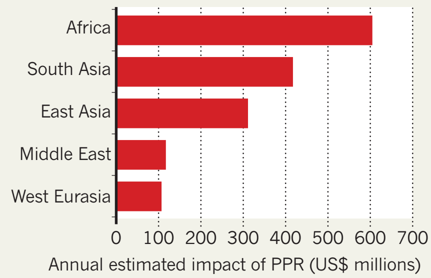
One positive effect of the campaign will be the construction of veterinary infrastructure in these areas. This will have impacts beyond PPR, for example by helping to combat other small ruminant diseases, such as goat and sheep pox. "There should be good knock-on effects for poor people," says Baron.

How exactly the eradication effort will be coordinated has yet to be hashed out. The UN Food and Agriculture Organization, which co-organized the Abidjan meeting, and the OIE are looking for Western governments, non-governmental organizations and charities to foot much of the estimated cost of $7.6 billion to $9.1 billion.

Samuel Thevasagayam, deputy director for livestock at the Bill & Melinda Gates Foundation in Seattle, Washington, which has contributed funds to PPR vaccine efforts in the past, says that the organization is evaluating whether to support the campaign. "It's a huge commitment — that's what causes donors to think carefully," says Mariner. He hopes that the management structure that emerges will be nimble and open to new ideas and approaches. "With PPR, we're going to have to continue to innovate," he says. ■

COST OF A GOAT-KILLER

The virus peste des petits ruminants, or PPR, costs farmers in lost production and livestock, and in caring for or vaccinating goats and sheep.



SOURCE: OIE/FAO

CLIMATE SCIENCE

Physicists, your planet needs you

Climatologists highlight cloud mysteries in an attempt to lure physicists to their field.

BY QUIRIN SCHIERMEIER

Climate science needs more mathematicians and physicists. So say prominent climatologists who are trying to spark enthusiasm for their field in budding researchers who might otherwise choose astrophysics or cosmology. Talented physical scientists are needed to help resolve mysteries that are crucial to modelling the climate — and, potentially, saving the planet — the group says, such as the ways in which clouds are formed.

There is a misconception that the major challenges in physical climate science are settled. "That's absolutely not true," says Sandrine Bony, a climate researcher at the Laboratory of Dynamic Meteorology in Paris. "In fact, essential physical aspects of climate change are poorly understood."

To attract physics and mathematics students to the speciality, Bony and her collaborators have presented some of the field's grand challenges in magazines such as *Physics Today* (B. Stevens and S. Bony *Phys. Today* http://doi.org/3f9; 2013), and are organizing summer schools for students from an array of scientific backgrounds.

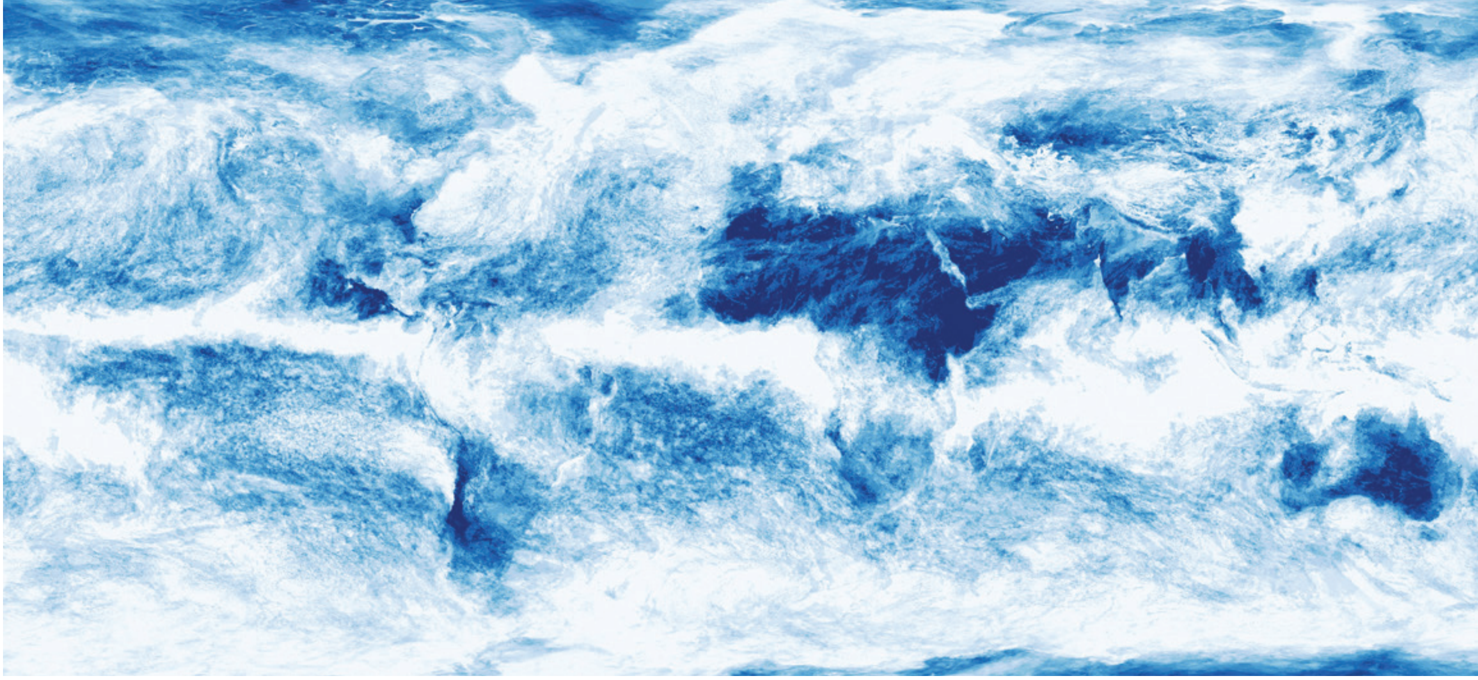
Last week in *Nature Geoscience*, Bony's team outlined four of the field's deepest questions, including how clouds and climate interact and how the position of tropical rain belts and mid-latitude storm tracks might change in a warming world (S. Bony *et al. Nature Geosci.* http://doi.org/3gb; 2015). The questions are best tackled, says Bony, by creating more realistic climate simulations — an approach that she hopes will appeal to physicists.

The perception that climate science is 'solved' is an inadvertent result of pressure on climatologists to convey a simple message to the public — for instance, that all dry regions will get dryer and all wet regions wetter in a warming climate, says Piers Forster, a climate modeller at the University of Leeds, UK. That has made the science "sound somewhat dull", he says.

"We too quickly turn to the policy implications of our work and forget the basic science," adds Bjorn Stevens, a director at the Max Planck Institute for Meteorology in Hamburg, Germany, and a co-author of the *Nature Geoscience* paper. Although climate scientists agree on the basics — for example, climate change is primarily the result of human activity — large uncertainties persist in 'climate sensitivity', the increase in average global temperature caused by a given rise in the concentration of carbon dioxide.

As Bony and co-authors argue, understanding how the warming climate might affect cloud cover, which influences the amount of sunlight reflected back into space and thus Earth's energy cycle, is key to addressing these uncertainties. A major weakness of current climate models is their limited ability to simulate the convection by which humid air is lifted into the atmosphere and which drives cloud formation and rainfall. In some instances, the models cannot even agree on whether the future will bring more rain or less.

Building better cloud-resolving models requires enormous computer power, as well as people who have a deep understanding of



MODIS ATMOS. SCI. TEAM/RETO STOCKLI, NASA EARTH OBSERV.

Clouds are key to understanding climate change, but more-realistic models of their formation are needed.

climate physics combined with skills in numerical modelling. But the number of scientists involved in developing computer algorithms for improved climate models is tiny, says Christian Jakob, an atmosphere researcher at Monash University in Clayton, Australia.

Physicists agree that climate science is not a big attractor of physics students. "Very few, and rarely the best, choose to do a master thesis in climatology," says Thierry Fichefet, a physicist and climate modeller at the French-speaking Catholic University of Leuven in Belgium. "Talented physicists commonly go into more glamorous fields such as astronomy, cosmology or particle physics." According to the American Institute of Physics in College Park, Maryland, 49 PhDs were awarded in atmospheric chemistry and climatology in the United States in 2013, compared with 303 for astronomy and almost 2,000 each for physics and mathematics.

Many physicists applaud Bony's effort to raise interest in climate science, but whether physics students will heed her call remains to be seen. "We offer courses in climate science and our students do recognize the importance of the field," says Paul Linden, a fluid-dynamics researcher at the University of Cambridge, UK. However, he says, classical subjects with a long history such as cosmology, are just more attractive, particularly at his university. "Most physics students would rather study with someone like Stephen Hawking, who is a member of our faculty." ■

BIOTECHNOLOGY

Synthetic biology called to order

Meeting launches effort to develop standards for fast-moving field.

BY ERIKA CHECK HAYDEN

Synthetic biologists have a vision. Researchers in this young field, who build 'devices' from engineered genes and other molecular components, imagine a future in which products such as drugs, chemicals, fuels and food are manufactured by microbes. These devices could even be wired up to create cellular computers, much as electronic transistors are wired up to make microprocessors (see *Nature* http://doi.org/3fz; 2013).

But if the dream is to be realized, those components need to perform more consistently and be more reproducible than they are now, especially as they move from the lab bench to the biofactory. Unlike silicon-based electronic devices, synthetic organisms assembled from genetic components do not always have predictable properties — at least not yet.

On 31 March, representatives from industry, academic institutions and government met at Stanford University in California to launch the Synthetic Biology Standards Consortium, an initiative led by the US National Institute of Standards and Technology (NIST) to address issues preventing the field from reaching its potential.

"It's the signal of a maturing industry," says Patrick Boyle, who oversees the organism-design pipeline at Ginkgo BioWorks, a synthetic-biology company in Boston, Massachusetts. "As we get better at synthetic biology, we want to make sure we are comparing apples to apples."

The standards push comes at a pivotal point for synthetic biology. Ginkgo BioWorks is one of several 'foundries' set up to mass produce organisms for various uses. Others include Amyris Biotechnologies and Zymergen, both in Emeryville, California, and publicly funded initiatives in the United Kingdom and the United States. Massive firms — and potential customers for these foundries — are showing interest in the field. The chemical manufacturers Dow and DuPont, the US defence-technology giant Lockheed Martin, the drug maker Novartis and the Dutch health and materials-sciences company DSM were all represented at the NIST meeting.

Participants divided into work groups to brainstorm what standards would make it easier for synthetic biologists to share methods, materials and information. The groups concentrated on a wide range of topics, including standards for automating methods, describing and assembling components and documenting the performance of engineered bacterial strains. One group considered how to demonstrate the safety of commercial synthetic-biology products. Another worked towards calibration methods for flow cytometry, a widely used technology for counting and sorting cells.

Each working group will now carry its ideas forward. NIST will provide support for these efforts, but where they go is up to the researchers. ▸

▶ Like most life-sciences fields, synthetic biology faces issues with reproducibility (see *Nature* **515,** 7; 2014). It has proved difficult for labs to replicate strains engineered by others, which is hampering progress in industry and in academic research. One lab may not provide enough information about how a part was made for another to reproduce the work. Or researchers may toil to refine a process, only to find that the developer has already done that work without reporting it in the literature.

"It's just difficult to know all the information out there, because there isn't a set of standard ways to describe what we know about a biological resource," says Ryan Ritterson, a synthetic biologist at the University of California, San Francisco.

Synthetic biologists who work in industry pushed for standards that would simplify commercial decisions. For instance, organisms that make products well in the lab do not always work when grown to bigger scales. Boyle advocated the development of 'reference strains' whose behaviour has been characterized in different types of fermentation equipment and growth media. This is particularly important to companies that want to understand whether variations in manufacturing efficiency stem from the organisms or from external factors.

The NIST-led effort is not the first attempt at standardizing the tools and methods of synthetic biology. The Synthetic Biology Open Language project is an online consortium that is developing standard nomenclature, symbols and other tools to describe engineered systems. The BioBricks Foundation has designed a licence to facilitate the free exchange of biological parts. And several repositories make and distribute such parts.

But those efforts have addressed only some of the reproducibility issues in synthetic biology. Parts in some of the repositories do not always work or are mischaracterized. Not all companies want to use parts that are in the public domain, or contribute their components to these repositories. And the field has moved on since some of the standards were set; for instance, the standard method for assembling BioBricks can be slower than newer methods for making complex DNA assemblies.

Meeting participants said that they hoped that the NIST initiative's inclusion of researchers and companies would help it to overcome some of those problems. "We had people coming from different areas who all have different stakes in the outcome," Ritterson says, "and rather than dividing into factions and deciding what standards would work best in our applications, we had thoughtful conversations about the standards that would work best for the entire community." ■

Tom Kariuki will head a funding platform for African research that is due to be launched in June.

SVEN TORFINN

DEVELOPING NATIONS

Africa aims for research autonomy

Regional hub intends to manage international grants and develop science strategy.

BY LINDA NORDLING

African scientists look set to gain greater control over research in their own countries, if an ambitious plan for a regional hub to award grants and develop research capacity bears fruit.

Three international funding bodies are giving seed cash of around US$4.5 million to establish the Alliance for Accelerating Excellence in Science in Africa (AESA). The London-based biomedical charity the Wellcome Trust also hopes to transfer the management of millions of dollars in its research funds to the alliance. AESA's other two backers are the UK Department for International Development and the Bill & Melinda Gates Foundation in Seattle, Washington. The idea is that AESA will be a platform for managing Africa-focused research programmes and a think tank to direct the continent's science.

"Science can and will transform Africa. But to get there, we must train critical numbers of excellent scientists in all corners of Africa. That is the mission of AESA," says Tom Kariuki, a Kenyan immunologist who was appointed as the alliance's director in March. It is due to be launched in June by African heads of state, and will operate out of the headquarters of the African Academy of Sciences in Nairobi.

REMOTE CONTROL

For decades, African science capacity and research output have lagged behind those of the rest of the world. But they are now taking off in fields with clear impacts on African development, such as health and agriculture, in nations including Uganda, Kenya, Ghana and Nigeria (see *Nature* **474,** 556–559; 2011). One problem is that overseas funders still supply a large chunk of the research cash and decide where and how it is spent.

"Much of the research done in Africa is still predominately financed by global funders from Western Europe and the United States, and still managed from Western capitals from funders' head offices," says Kariuki (see 'Funding from abroad'). That has limited the impact of such research, in part because it matches priorities set outside Africa. Funding is in short supply for studying neglected tropical diseases, for example, and funding for HIV research is not always directed at the countries in the greatest need. African researchers can also struggle to keep

teams together once overseas grants run out.

"It's weird that for 40 years, the agenda-setting and the funding decisions for research in Africa has been done from London, Seattle, Geneva or wherever," agrees Kevin Marsh, a clinical epidemiologist at the University of Oxford, UK, and a senior adviser on the AESA initiative.

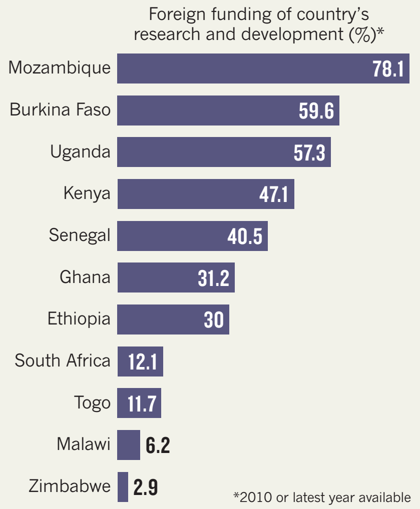
Instead, AESA will invite funders both on and outside the continent to delegate the peer-review and grant management of their African programmes to the alliance. The idea is to shift the centre of gravity for African funding decisions to the continent, says Simon Kay, head of international operations at the Wellcome Trust. AESA wants to create more buy-in from African governments on the research being done, Kay adds.

MONEY MANAGEMENT

As a start, the Wellcome Trust is considering handing over the management of its five-year, £40-million (US$60-million) Developing Excellence in Leadership, Training and Science initiative to AESA later this year. This programme, launched last September, aims to build up research capacity and train leaders who can drive regional agendas, by awarding competitive grants, initially in health research. It expects to announce the winners of its first crop of applications in May. The charity will cede more control only if it is sure that Nairobi can manage the programme to its own standards — so AESA staff will undergo a year of training.

FUNDING FROM ABROAD

Many African nations depend on foreign grant money (often from overseas organizations) to fund their research.



SOURCE: NEPAD

Marsh says that other funders have signalled their eagerness to hand over programmes to AESA. "Let's say we start with £40 million. I'd be disappointed if in a year's time we haven't at least doubled that. And in the long term, we have to move to hundreds of millions," he says.

The funding cannot come just from international donors, African scientists agree. "This is a great initiative. But it will be stillborn unless African governments put money into it," says Salim Abdool Karim, a clinical epidemiologist and director of the Durban-based Centre for the AIDS Programme of Research in South Africa. The hope is that AESA would be attractive because it offers governments a way of awarding merit-based science grants without having to train their own grant managers and set up research funders nationally.

A smaller partnership has already been attempted by the European Union: from 2011 to 2013, it gave €14 million (US$15 million) to the African Union in Addis Ababa to manage competitive grant calls in research areas including agriculture, water and sanitation. But African governments have not followed up on that effort by chipping in with their own money.

AESA has not yet secured any African national government funding. But it will receive $500,000 towards its setting up from the New Partnership for Africa's Development, a continental body for making and implementing policy, with headquarters in Pretoria. Kariuki says that AESA will also lobby African governments to support research in their own countries. ■

THE FUTURE OF THE POSTDOC

There is a growing number of postdocs and few places in academia for them to go. But change could be on the way.

BY KENDALL POWELL

By the time Sophie Thuault-Restituito reached her twelfth year as a postdoctoral fellow, she had finally had enough. She had completed her first postdoc in London, then moved to New York University (NYU) in 2004 to start a second. Eight years and two laboratories later, she was still there and still effectively a postdoc, precariously dependent on outside grants to secure and pay for her position. Her research on Alzheimer's disease was not making it into high-profile journals, so she was unable to compete for academic positions in the United States or Europe. She loved science and had immense experience, but with two young children at home, she knew she needed something more secure. "My motivation was gone. I was done with doing research," she says.

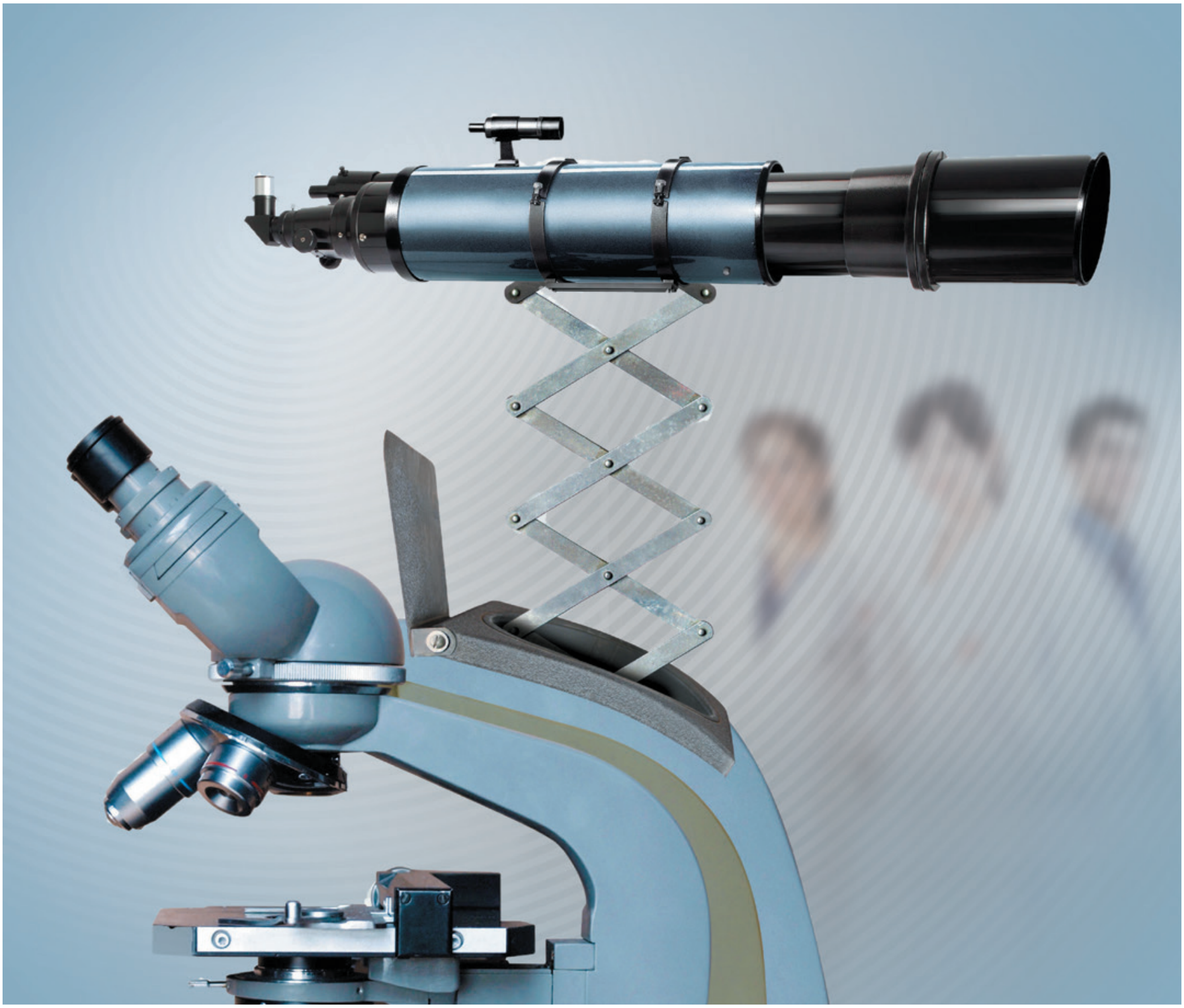
So in 2013, Thuault-Restituito moved into a job as a research-laboratory operations manager at NYU, where she coordinates building renovations and fosters collaboration between labs. She enjoys the fact that her staff position has set hours, as well as better pay and benefits. But at the time of the move, she mourned the loss of a research career and she regrets the years wasted pursuing one. "I stayed five years more than I should have," she says.

Thuault-Restituito is the face of a postdoctoral system that is broken. These highly skilled scientists are a major engine driving scientific research, yet they are often poorly rewarded and have no way to progress in academia. The number of postdocs in science has ballooned: in the United States alone, it jumped by 150% between 2000 and 2012. But the number of tenured and other full-time faculty positions has plateaued and, in some places, it is even shrinking (see *Nature* **472,** 276–279; 2011). Many postdocs move on to fulfilling careers elsewhere, but those who want to continue in research can find themselves thwarted. They end up trapped as 'permadocs': doing multiple postdoc terms, staying in these positions for many years and, in a small but significant proportion, never leaving them. Of the more than 40,000 US postdocs in 2013, almost 4,000 had been so for more than 6 years (see 'The postdoc pile-up').

This problem is felt acutely in the large US biomedical-sciences workforce, but the trends are similar in many other countries and disciplines — and the economic drivers are too. Postdoc salaries have remained low — often less than the stipend and tuition costs of a graduate student. "We had the incentives all wrong," says Paula Stephan, an economist at Georgia State University in Atlanta who studies research labour markets. "We made postdocs so cheap that principal investigators had lots of incentives to hire them."

Discussion about the postdoc problem has grown increasingly loud. In December 2014, a committee convened by the US National Academies released a report aimed at highlighting and improving the postdoc's plight. The

ILLUSTRATION BY VIKTOR KOEN



committee called for a hike in salaries, from the current recommended starting salary of US$42,840 to $50,000, and a 5-year limit on the length of postdocs. Senior scientists in the United States, who have been urging reforms for the scientific workforce as a whole, have identified the postdoc oversupply as one of the most urgent issues (B. Alberts *et al. Proc. Natl Acad. Sci. USA* **112,** 1912–1913; 2015).

Experts acknowledge that change will be hard; after all, the National Academies made similar recommendations 15 years ago with little effect. But some institutions and countries have started to address the issue. Several US universities have enforced 5-year term limits, New Zealand inadvertently narrowed the pipeline when it slashed the number of postdocs available, and some laboratories are moving permadocs into stable, better-paid positions. Other scientists who are keen to help postdocs are watching the results with interest. "We've always been at risk of producing more scientists than we have places for, but the stresses and strains were not harmful in the way they are now," says Shirley Tilghman, president emerita of Princeton University in New Jersey, who has studied the workforce problem. "Some changes will have to happen."

THE FIXED-TERM POSTDOC

In 2008, while Thuault-Restituito was there, NYU's School of Medicine decided to try a tough-love approach: it began enforcing a rule that researchers could hold a postdoc for a maximum of 5 years — including time spent at other institutions. In 2014, 35 of the roughly 400 postdocs there left because their time was up.

The time limit can be painful for people who feel forced out, says Keith Micoli, chairman of the board of the National Postdoctoral Association and director of the NYU School of Medicine postdoctoral programme. "People coming up against it put me in an ethical quandary: what's best for that postdoc and what's best for postdocs as a whole?"

Micoli says that term limits combat two problematic phenomena. The first is the 'just one more year, experiment or paper' syndrome, in which postdocs feel that they must endlessly build their academic CV before moving on. The second is the permadoc who stays on indefinitely, eventually runs into his or her adviser's retirement and is stranded without a job, a situation that Micoli himself encountered. Having a hard deadline forces postdocs to make career decisions and "people are better for it", he says. Of the postdocs who left NYU in 2014, Micoli says that roughly equal numbers got faculty positions and left academia.

Other major research universities, such as the University of California system and the University of North Carolina, Chapel Hill (UNC), have also implemented 5-year term limits. But the limits are not always strictly enforced. Postdocs and their advisers can

often request a sixth year and some postdocs are moved into positions that are postdocs in everything but name. When Thuault-Restituito bumped up against the 5-year rule in 2006 — before the university enforced it more strictly — she was promoted to 'associate research scientist', a staff position that brought better benefits, but no extra pay or job security. Her position was still contingent on outside grant funding, which was far from guaranteed. "At the end of the day, my job and what I was doing in the lab didn't change at all," she says.

Sibby Anderson Thompkins, who directs the postdoctoral affairs office at UNC, says that the most-recent postdocs there embrace the term limit. They enter with a plan to find a career path quickly and exit the postdoc early if an opportunity arises. Anderson Thompkins, who also sat on the 2014 National Academies report committee, says that this type of planning should begin in graduate school, alongside raised awareness of the academic bottleneck that trainees will face. Whereas about 65% of US PhD-holders continue into a postdoc, only 15–20% of those move into tenure-track academic posts. The European situation is even more competitive — in the United Kingdom, for example, about 3.5% of science doctorates become permanent research staff at universities.

Term limits have also been tested in the United Kingdom, France and Germany, where labour laws limit the number of years that academic researchers can remain on short-term contracts before they must be hired permanently. But it is unclear whether these laws help or hurt, because there are often ways around them.

In Germany, for example, a law originally intended to curb postdoc contracts to about six years after completing a PhD was altered so that scientists can remain on short-term contracts as long as they are funded by an external grant and not paid directly by the university. The result is that scientists surf endlessly from one postdoc to another: "There are unlimited numbers of short-term contracts," says Sibylle Anderl, a German postdoc in astronomy at the Grenoble Institute for Planetary Sciences and Astrophysics in France. "The real problem for German postdocs is that we don't have enough permanent positions available."

THE ELITE POSTDOC

Postdocs don't have to be forced out of the pipeline if, instead, they are never let in. That was the result when, in 2010, the New Zealand government decided to axe a scheme that had funded roughly 90 postdoc slots — eliminating nearly one-third of its postdocs in one fell swoop.

Before this, the government covered salaries for a huge chunk of the country's postdocs, who enjoy salaries and benefits nearly equivalent to those starting permanent academic positions. For most labs, postdocs are too expensive to fund from research grants. So when the government funding disappeared — mainly a money-saving decision — so too did many postdoc spots.

Lara Shepherd got caught in the squeeze when fellowships vanished in her field of evolutionary biology and she reached the end of her first postdoc, at the Museum of New Zealand Te Papa Tongarewa in Wellington. She secured a second postdoc at Massey University in Palmerston North, using a grant to pay half of her salary and working part time to cover the rest. But she could not land a coveted academic position. "New Zealand is so small — there are very few jobs in your particular area of expertise," she says.

"I THINK THE GOAL IS TO MAKE THE POSTDOC SOMETHING SPECIAL. IT SHOULD BE HARD TO GET A POSTDOC — HARDER THAN GETTING INTO GRADUATE SCHOOL."

Shepherd eventually found a temporary research position back at the Museum of New Zealand, and scored an early-career grant from the Royal Society of New Zealand, which she leveraged into a permanent position. She now oversees genetic analyses of plant, animal and fossil samples. Without the early-career fellowship, she says, "I would have been looking outside of science."

Many principal investigators (PIs) in New Zealand are unhappy with the situation. With no postdocs to help them, they struggle with lab management and mentoring, and they say that labs have become dependent on graduate students. "All we've done is to outsource our postdocs," says Shaun Hendy, a physicist at the University of Auckland. "We've removed a cohort of young researchers from our system and replaced them with even younger, less-experienced researchers."

Once trained, the country's best PhD students tend to head out of science or to postdocs overseas. One lab head describes a top marine-biology graduate who — with no prospect of a postdoc or academic job — ended up driving a forklift before eventually landing a position in the country's statistics bureau. Hendy predicts that the postdoc void will result in lower-quality, less-complex research projects. "I'm sure there will be productivity hits down the line."

Simon Davy, head of the school of biological sciences at the Victoria University of Wellington, says that the research culture of university departments loses vibrancy without any postdocs. His department of 35 research groups hosts fewer than 10 postdocs. His own lab has been lucky enough to have a couple of them in the past 5–6 years and he says that this has tripled his group's productivity.

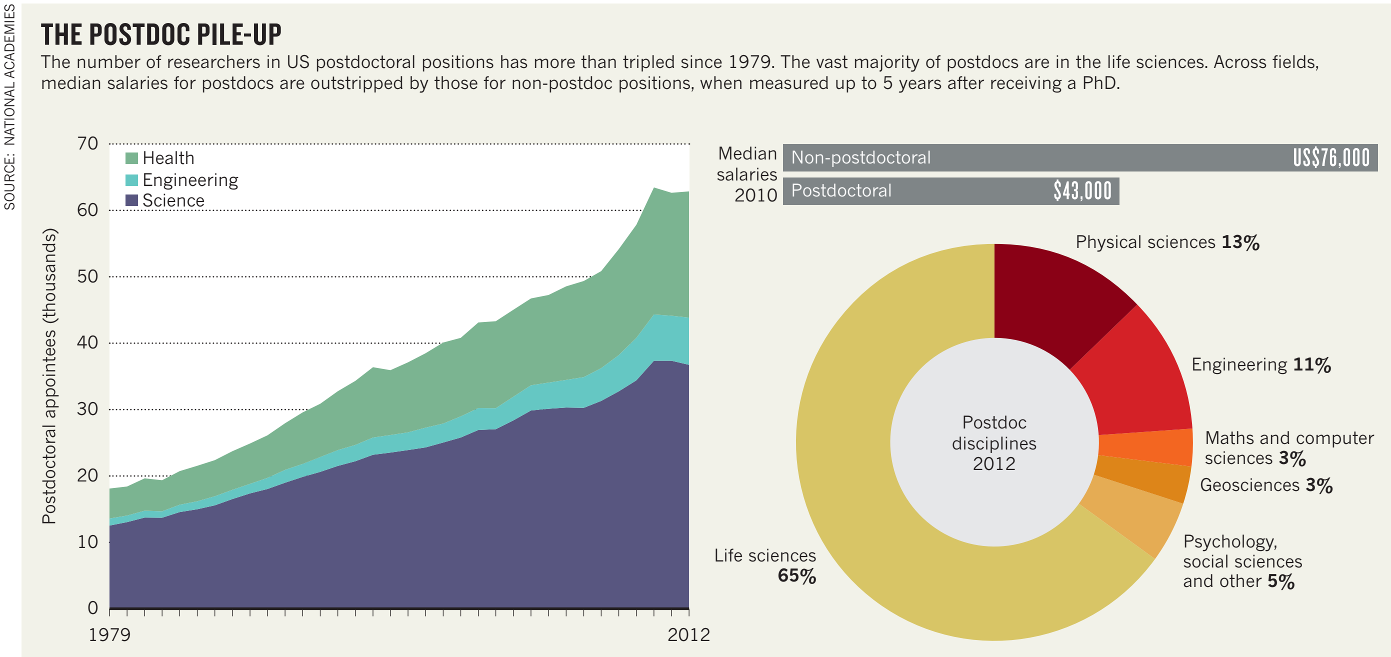
If Davy could wave a magic wand and bring back the government-funded postdoctoral positions, he would — and so would 560 of the country's scientists, who, in 2011, collectively sent a letter of protest to the science minister, among other government leaders. "I've struggled to think of positives from our experience," Hendy says. Science-development manager Anne Berryman, from the New Zealand Ministry of Business, Innovation and Employment, says the decision to cut postdocs was designed to reprioritize government support towards later career stages, and contends that there is no evidence of harm to the country's scientific research.

Most US researchers balk at the idea of restricting the number of postdocs entering the system. Jennifer Lippincott-Schwartz, a cell biologist at the US National Institute of Child Health and Human Development in Bethesda, Maryland, says that it is nearly impossible to determine who has the characteristics of a superstar researcher until mid-way through a postdoc term. "I don't think it's bad when part of that workforce has to leave and move into other professions," she says. "They carry with them skills that are not wasted. They still have a knowledge base that is valuable to society."

THE SUPERDOC

If postdocs are so prized, then one obvious solution is to reward them. Both the 2014 National Academies report and earlier reports urged US lab heads to consider creating senior staff scientist, or 'superdoc', positions. These would be higher-paid, permanent jobs for talented postdocs who have no desire to start their own labs.

Some funding agencies and institutions around the world already offer this option. Lippincott-Schwartz, for example, has two superdocs in her cell-biology laboratory at the National Institutes of Health (NIH). One serves as a software developer for the lab's super-resolution imaging of intracellular structures. The other is a microscopy specialist and lab manager. They both mentor trainees, help to write publications and keep up with the latest technological advances in the field. "These staff scientists offer so much to individual laboratories," she says. "They can do the science they love without dealing with all the



bureaucratic stuff associated with being the PI." Each of her superdocs earns $20,000–30,000 more than postdocs typically earn — a cost she was able to cover by requesting more funds for her lab's annual budget from the NIH.

But other lab heads say that they struggle to find the resources to pay for superdocs, and without an increase in funding, the inevitable trade-off is fewer workers. That reality is hard to stomach for lab heads who are trying to balance the pressure to produce results and papers — generally maximized by lots of staff on low salaries — with the desire to keep and promote experienced employees. "It's economics, and we need to face up to that. There may not be as many people working in your lab. No one wants to talk about that," says Micoli.

One scientist struggling with this dilemma is Leslie Leinwand, a molecular cell biologist at the University of Colorado Boulder's BioFrontiers Institute. She relies on two postdocs, Massimo Buvoli and Steve Langer, who have been in her lab for nearly two decades. But if she created staff-scientist positions for them — as the National Academies report recommends — the two increased salaries would equal nearly two-thirds of the annual budget for a typical NIH R01 grant, on which many biomedical labs rely. "There needs to be a place for such people who just want to stay at the bench, but I stay awake at night worrying about salaries for Massimo and Steve. Frankly, I can't afford to pay them what they deserve," Leinwand says. Anne Carpenter, a computational biologist at the Broad Institute in Cambridge, Massachusetts, requested extra grant funds to hire more permanent scientists rather than trainees, but found that her proposals were criticized by grant reviewers, who questioned why she was using such expensive staff to do the work.

Some funding bodies do offer funds specifically for staff scientists, and others are introducing them. In March, the US National Cancer Institute proposed a grant programme designed for superdocs that would cover a salary in the range of $75,000–100,000 for five years. It is planning to grant 50–60 such 'research specialist awards' throughout the next 18 months (see page 255).

THE REINVENTED LAB

The real solution to the postdoc problem, Tilghman says, lies in dramatically changing the composition of labs to make them smaller, with a higher ratio of permanent staff scientists to trainees. This was also a key recommendation in the National Academies report. "The more I have thought about this question, the more I'm convinced that at the heart of the problem is the structure of the lab," says Tilghman, who headed up a 2012 study of the NIH workforce (see go.nature.com/wsqzgj).

The biggest challenge, she says, is persuading lab heads to embrace such a model when there is a tremendous bias in favour of the cheap labour that graduate students and postdocs represent. But that bias is short-sighted, she argues, when one staff scientist can do the work of three less-experienced researchers. "We've got to persuade faculty that this is a true trade-off, and a positive trade-off for their research productivity."

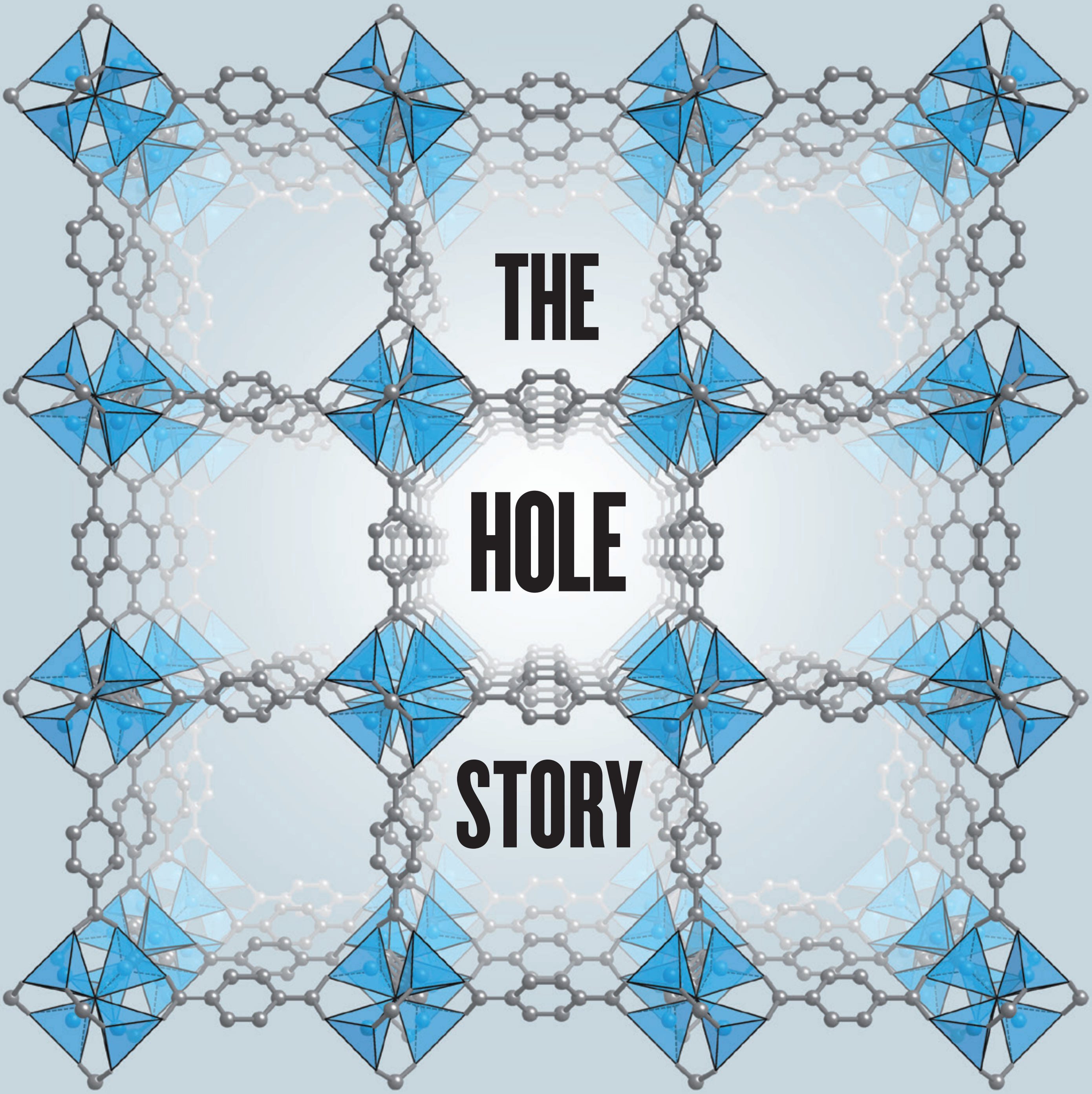
Labs stuffed full of trainees do not always translate to better results, says Gregory Petsko, chair of last year's National Academies committee and a neuroscientist at Weill Cornell Medical College in New York City. "I don't think many of us need the labs to be the size we have them." Petsko proposes combining various strategies — term limits, fewer postdoc positions and more staff scientists — to deflate the swollen postdoc population. That would stop the postdoctoral fellowship from being the default step after earning a doctorate. "I think the goal is to make the postdoc something special," he says. "It should be hard to get a postdoc — harder than getting into graduate school."

The question is, can the scientific community be convinced? No one interviewed for this story — whether lab heads or postdocs themselves — wanted to give up these highly valued research positions. But few lab leaders, institutions or funders seem willing or able to spend what it takes to reward them appropriately. Petsko says that funding agencies could step in and enforce change, by demanding that universities direct a portion of their overhead payments — money given to the university rather than the lab — towards creating more staff-scientist positions.

Davy points out that the solution needs to be global, or else postdocs denied jobs in one country will simply slide across country borders to find them elsewhere. In an ideal world, he says, postdocs would be able to take their funding wherever they like. "People should be going to the best labs, the best places for them to work and be trained, which are dotted around the world."

As for Thuault-Restituito, she does not regret her postdocs. But if she had to walk that path again, she would move into another career much earlier. She agrees that fewer PhDs should be flowing into postdocs, and is frank with graduate students who ask her for advice: "If you are not 150% sure you want to do it right now, don't do a postdoc." ■

Kendall Powell *is a freelance writer based in Lafayette, Colorado.*



THE HOLE STORY

Swiss-cheese-like materials called metal–organic frameworks have long promised to improve gas storage, separation and catalysis. Now they are coming of age.

BY MARK PEPLOW

Sprawled across a vast site on the river Rhine in Germany is a small city built from glittering steel: the headquarters of chemical giant BASF. Boasting a daytime population of about 50,000 people, it is criss-crossed by a grid of streets bearing names that commemorate the company's stock in trade: Methanolstrasse, Ammoniakstrasse, Gasstrasse.

Over the past two years, a small fleet of delivery vans and cars has clocked up thousands of kilometres on these streets while carrying a big secret: fuel tanks packed with an unusual crystalline material that is riddled with pores roughly a nanometre wide. Inside these pores sit methane molecules arrayed in neat stacks, ready to fuel the vans' combustion engines.

The crystals are metal–organic frameworks (MOFs), molecular scaffolds made up of metal-containing nodes linked by carbon-based struts

YAGHI LABORATORY

(see 'An open box'). The resulting pores are ideal for trapping guest molecules and, in some cases, forcing them to participate in chemical reactions. And they can be tailored with exquisite precision: researchers have created more than 20,000 types of MOF, with potential applications that range from stripping carbon dioxide from power-plant exhausts to separating intractable industrial mixtures, catalysing chemical reactions and revealing molecular structures. "MOFs are the fastest growing class of materials in chemistry today," says Omar Yaghi, a chemist at the University of California, Berkeley, and one of the pioneers of the field.

MOFs were long thought to be too frail for use in the real world, often collapsing as soon as the guest molecules were removed. Many researchers were sceptical that the products could ever compete against the tough inorganic materials called zeolites, whose pores are exploited in a wide variety of industrial processes, including filtration and catalysis.

But after more than a decade of intensive research in labs around the world, MOFs are poised to make their debut in commercial applications. Although unwilling to reveal the identity of the MOF in question, BASF has said that it is ready to market a methane-storage system this year that can cram in much more fuel than a conventional pressure vessel.

MOF researchers say that this milestone would be a shot in the arm for their work, and potentially help to stimulate commercial interest in the many other applications that are close behind, often from other producers.

"I thought, 'Wow! This is a way to revolutionize the way organic chemistry can progress.'"

STORAGE WARS

Much of the ferment in MOFs dates back to 1999 and the debut of two unusually robust varieties: HKUST-1, developed at the Hong Kong University of Science and Technology[1]; and MOF-5, developed by Yaghi[2]. The latter has an internal surface area of at least 2,300 square metres per gram — enough to cover more than eight tennis courts. "That was the turning point, because it broke all surface-area records," says Yaghi. "Years later, BASF told me that they thought it was a misprint."

More internal surface area means more places to stack guest molecules, and Ulrich Müller, who leads BASF's research on porous materials, was quick to see an opportunity. "We started working on MOFs directly after Yaghi's paper," he says, and the pair quickly forged a collaboration that continues to this day.

The key to making stable MOFs is to use clusters of metal atoms as the nodes, rather than individual ions. The geometry of the clusters determines the overall architecture of the crystal, which can be held together by a cornucopia of organic linkers. The growing set of interchangeable Tinkertoy components makes MOFs much more adaptable than zeolites and enables chemists to design products with pores that have just the right size and chemical properties for specific applications. Today, there are MOFs that can withstand temperatures of 500 °C, or easily endure a week in boiling methanol; others have internal surface areas that are triple that of MOF-5, or pores large enough to accommodate chunky proteins[3].

BASF currently dominates the nascent MOF market. It has targeted methane storage because shale gas is cheap and increasingly available, so could be used to power automobiles that incur lower running costs and generate less CO_2 than conventional vehicles. However, at present, the gas needs to be stored in bulky and expensive high-pressure tanks, which is a major disincentive. MOFs could overcome that by storing more methane at lower pressures.

To make that application work, the size and chemical properties of the MOF's pores must be just right, because they determine how the methane molecules stack up inside. "If you just have methane floating inside the pore, you might as well use an empty canister," says Yaghi.

To tie down the methane, researchers use MOFs with pores that boast exposed metal ions. The ions distort methane's electron cloud, polarizing it so that the gas molecules stick to the metal. But if the pores bind to methane too weakly, the gas will leak out; too strongly, and the vessel will be hard to empty. The best MOF crystals occupy a Goldilocks zone that gives them at least twice the capacity of an empty vessel at moderate pressure, yet still allows them to release almost all their methane as the pressure drops. "Methane storage for automobiles is largely a solved problem," says Yaghi.

➲ NATURE.COM
To hear more about MOFs, visit.
go.nature.com/maemh5

Commercial success is far from guaranteed, however. Since the price of crude oil began to plummet last year, the economic incentive of gas has disappeared. "This gap is currently almost non-existent," says Müller. "Everything is in a little bit of turmoil because of that."

Market-watchers predict that the price of oil will rebound sooner or later. But in the meantime, Jeffrey Long at the University of California, Berkeley, says that there is ample scope to improve MOF methane-storage systems. In collaboration with Yaghi, BASF and the Ford Motor Company, he is aiming to reduce the pressure needed to fill a tank. "If you go down to 35 bar, people can potentially fill their cars at home," he says. Long and his colleagues say that they have created a MOF that stores more methane than the best current compounds at low pressure and are preparing to publish their results. "We can beat it by a reasonable margin," Long says.

MOFs could make an even bigger impact on transportation by storing hydrogen for fuel-cell vehicles. Compressing chilled gas into high-pressure tanks is complex and expensive. But replacing those tanks with MOFs that can store useful amounts of hydrogen is a tough challenge. "There's no absorbent out there that has a high enough capacity to be used commercially," says Long.

Long's team has developed[4] a record-breaking nickel-based MOF that binds to hydrogen strongly enough to carry 12.5 grams of the gas per litre of storage tank at room temperature and 100 bar. That is, however, well short of the US Department of Energy's hydrogen-storage target for 2020, which calls for a corresponding figure of 40 grams per litre. Using MOFs with metal ions in their pores that can each bind to several hydrogen molecules could bring researchers closer to that goal.

In the meantime, others are looking to commercialize MOFs for niche gas-storage applications. Omar Farha at Northwestern University in Evanston, Illinois, co-founded the spin-out company NuMat Technologies in Skokie, Illinois, in 2012 to develop MOFs that can safely store some of the toxic gases used in the semiconductor industry, including boron trifluoride, phosphine and arsine. "We're doing something different from everybody else," he says. "It's a smaller market that we can capitalize on very quickly."

Farha reckons that the company's first product will launch in the next two years, helped by a recent boom in the use of computer modelling to predict the properties of MOFs. In 2012, Farha and his colleagues showed that they could reliably screen almost 140,000 hypothetical MOFs for their methane-storing ability[5], and now they are saving time and money by synthesizing only those MOFs that show promise in similar computational tests.

TRIAL SEPARATION

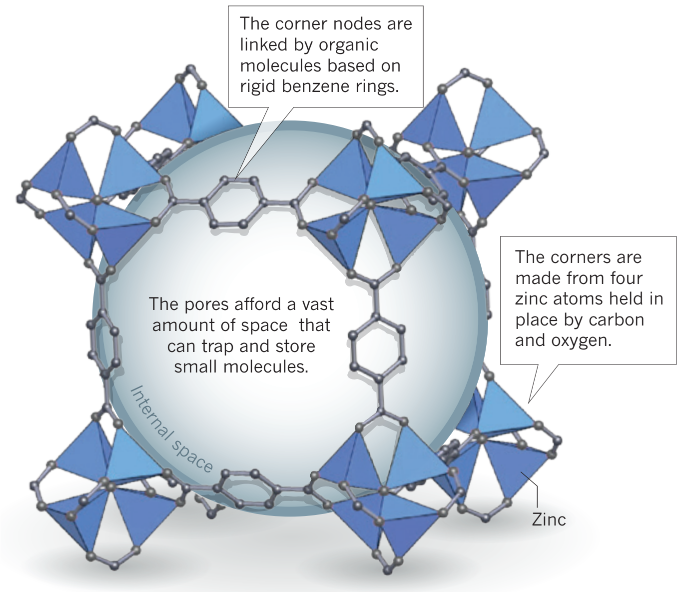
Researchers are also hoping that MOFs can pluck specific molecules out of the air — literally. "Gas separation, in particular, is where these materials could have a competitive advantage," says Long.

They could be particularly attractive for industrial cracking plants, which heat crude oil to break down large molecules into light hydrocarbons. These gases can be extremely difficult to separate. Propene and propane differ by just two hydrogen atoms, for example, and their boiling points are only around 5 °C apart. At the moment, refiners isolate them by cooling the mixture until it liquefies, then warming it slowly so that first one gas and then the other boils off. But these temperature swings make it one of the most energy-intensive processes in the chemical industry.

Long's group has shown that a crystal known as Fe-MOF-74 makes the job easier and potentially much less costly. The crystal's exposed metal cations can latch onto the electrons of a passing propene molecule, slowing down its passage. At a balmy 45 °C, propane emerges first; warming the MOF then frees a 99% pure stream of propene[6]. Another crystal, $Fe_2(BDP)_3$, can efficiently separate isomers of hexane[7], which comes in

AN OPEN BOX

One of the first metal–organic frameworks to be rugged enough for practical use was a crystal known as MOF-5. Its building blocks are cubes formed by linking together eight metallic nodes with carbon-based struts.



linear and variously branched forms. The linear molecules get stuck in the corners of the MOF's triangular channels, an architecture that Long says would be impossible to achieve with zeolites.

Perhaps the ultimate test of MOF-based separation would be to capture some of the 13.7 gigatonnes of CO_2 produced by fossil-fuel power plants each year. Conventional carbon-capture systems rely on solvents that react with the CO_2 in a power station's 40 °C exhaust stream. Removing and heating the solvent to 120 °C or more frees the gas for collection and storage. But swinging the temperature back and forth eats up 20–30% of the plant's power, and requires expensive infrastructure.

Last month, Long's team showed[8] that magnesium- and manganese-based MOFs could absorb and release more than 10% of their weight in CO_2 from flue gas with a temperature swing of just 50 °C. Their pores are lined with amine molecules that are similar to the solvents already used for carbon capture, and react with CO_2 to produce chains of tightly packed ammonium carbamate molecules.

A similar, as yet unpublished MOF can release its carbon cargo below 100 °C, and Long hopes to test it at the US National Carbon Capture Center in Wilsonville, Alabama. It has a higher working capacity and lower temperature swing than solvent systems, so Long expects to be able to reduce the size of the capture unit and cut infrastructure costs; he has already co-founded a start-up company, Mosaic Materials in Berkeley, to produce the MOFs.

CRYSTAL SPONGES

Industrial-scale development of any new material is slow work. But applications can blossom remarkably quickly if only small quantities are required. Just two years ago, Makoto Fujita at the University of Tokyo developed a MOF that can help to determine the structure of pharmaceuticals and other organic molecules. Now he is deluged with commercial interest.

Many organic molecules steadfastly refuse to form crystals, which normally rules out using the conventional technique of X-ray crystallography to determine the precise spatial arrangement of their atoms. But in 2013, Fujita's team showed that a zinc-based MOF could soak up molecules of miyakosyne A, a natural marine product, and hold them steady in its pores so that X-rays can reveal their structure[9].

"I thought, 'Wow! This is a way to revolutionize the way organic chemistry can progress,'" says Phil Baran, an organic chemist at the Scripps Research Institute in La Jolla, California.

Others, however, were less impressed. Crystallographers found the MOF's performance hard to reproduce, and then Fujita's team found an error[10] in its structure of miyakosyne A, making others wary of the technique. Since then, however, Fujita and others have produced detailed instructions[11] that are winning over sceptics. The technique cannot handle all molecules, but Fujita thinks that 20–30% of the organic compounds they test can have their X-ray crystal structure determined in this way, using as little as 5 nanograms of the guest molecule.

Last year, the Japan Science and Technology Agency awarded Fujita US$15 million over five years to help him to commercialize the method, and some pharmaceutical companies are now using the technique to assist drug development. And a Japanese reagent company plans to make Fujita's crystal sponge — and the successors being developed in his lab — available off-the-shelf within the next three years.

GO-FASTER MOFS

Catalysis has long been touted as one of the most promising applications for MOFs. Their tunable pores can hold reagents in place, cleave specific bonds and then forge new ones, just like the active site of an enzyme.

But until a few years ago, progress on such catalysts was very slow, says chemist Joseph Hupp at Northwestern University, not least because very few MOFs were chemically stable enough to put through multiple rounds of reactions[12]. As a result, says Hupp, "there is not yet an example of a reaction for which MOFs are so superior that a typical organic chemist would choose a MOF-based catalyst over an existing catalyst".

Now, however, researchers are making promising catalysts by taking stable MOFs and tweaking the chemical groups around their pores[3]. They have also gone one stage further, gradually swapping out entire linkers or metal nodes to transform the MOF's chemical and physical properties without collapsing the whole structure[13]. These advances have allowed chemists to design and make a much wider variety of rock-hard but chemically-active MOFs. "There's a lot of MOFs now that we just couldn't make five years ago," says Hupp.

Indeed, one of the growing challenges for the field is the bewilderingly large number of MOFs. "We have too many," says Yaghi. Hupp agrees. Researchers need to dial back on synthesizing MOFs whose properties are never fully explored, he says, and instead focus on refining those that have proven stability or activity.

Another challenge is that MOFs have to compete with incumbent technologies such as zeolites. This puts a premium on getting the cost down by making MOFs from abundant metals and cheap organic linkers that can be manufactured in safe and inexpensive processes. BASF, for example, is making MOFs at the tonne-scale in water, rather than in other solvents.

Yet MOFs can compete through their originality. Yaghi is developing MOFs that contain several types of pore within the same crystal, so that molecules would undergo a predefined sequence of reactions as they pass from one region to the next[14]. These MOFs could behave like microscopic versions of a chemical plant, allowing scientists to synthesize molecules piece by piece in a continuous process.

"That's our dream," says Yaghi, "and it's only possible in MOFs." ■

Mark Peplow *is a science journalist based in Cambridge, UK.*

1. Chui, S. S.-Y., Lo, S. M.-F., Charmant, J. P. H., Orpen, A. G. & Williams, I. D. *Science* **283,** 1148–1150 (1999).
2. Li, H., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. *Nature* **402,** 276–279 (1999).
3. Furukawa, H., Cordova, K. E., O'Keeffe, M. & Yaghi, O. M. *Science* **341,** 1230444 (2013).
4. DOE Hydrogen and Fuel Cells Program. *FY 2014 Annual Progress Report* (DOE, 2014).
5. Wilmer, C. E. *et al. Nature Chem.* **4,** 83–89 (2012).
6. Bloch, E. D. *et al. Science* **335,** 1606–1610 (2012).
7. Herm, Z. R. *et al. Science* **340,** 960–964 (2013).
8. McDonald, T. M. *et al. Nature* **519,** 303–308 (2015).
9. Inokuma, Y. *et al. Nature* **495,** 461–466 (2013).
10. Inokuma, Y. *et al. Nature* **501,** 262 (2013).
11. Inokuma, Y., Yoshioka, S., Ariyoshi, J., Arai, T. & Fujita, M. *Nature Protoc.* **9,** 246–252 (2014).
12. Lee, J. *et al. Chem. Soc. Rev.* **38,** 1450–1459 (2009).
13. Deria, P. *et al. Chem. Soc. Rev.* **43,** 5896–5912 (2014).
14. So, M. C. *et al. J. Am. Chem. Soc.* **135,** 15698–15701 (2013).

COMMENT

ILLUSTRATION BY PAUL BLOW

Reward bioinformaticians

Biological data will continue to pile up unless those who analyse them are recognized as creative collaborators in need of career paths, says **Jeffrey Chang**.

The US Precision Medicine Initiative, announced in January, relies on bioinformatics. The US$215-million project calls for collecting medical, physiological and genomic data from more than one million people in the United States, and aims to find patterns across individuals to improve health care. It does not address a worsening deficiency in the scientific community: biological data are accumulating faster than people's capacity to analyse them. For example, Sendurai Mani, a pioneer in cancer metastasis and genomics at the University of Texas MD Anderson Cancer Center, laments that he is constantly searching for bioinformaticians — those who analyse such data.

One explanation for the shortage seems obvious: there are not enough bioinformaticians. If so, the solution seems straightforward: train and hire more. As someone tasked with addressing this need at my institute, the University of Texas Health Science Center at Houston, I can vouch that the situation is not so simple.

The scientific community has failed to craft attractive career paths for those who do the analyses it increasingly requires. Institutions and funding bodies must carve out a viable place for bioinformaticians who focus on collaborations, and reward them for their abilities to navigate the myriad demands of multidisciplinary projects.

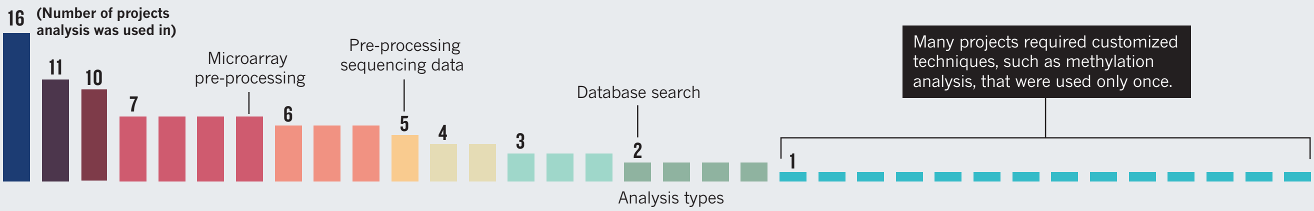
Biologists are increasingly finding that questions that are initially based on a single protein or gene quickly expand to require large-scale experiments. To support them, dozens of institutions have set up centralized bioinformatics facilities. Most were established in the past decade[1]; what many consider the first bioinformatics 'core' was set up about 20 years ago by Fran Lewitter at the Whitehead Institute in Cambridge, Massachusetts.

Until the past few years, my institute had limited bioinformatics expertise. In fact, grant reviewers deemed some applicants' data-analysis plans unlikely to succeed because of this. In part to address this, the university hired more faculty members, including myself, in 2010. We established our bioinformatics service centre in 2012.

To give greater support to researchers, our centre set out to develop a series of standardized services. We documented the ▶

ROUTINELY UNIQUE

Over 18 months, 46 data-analysis projects undertaken at the bioinformatics core of the University of Texas Health Science Center at Houston required 34 different types of analysis — most were used infrequently. Each project demanded unique combinations of analyses, demonstrating how bioinformaticians must be versatile, creative and collaborative.



SOURCE: UNIV. TEXAS HEALTH SCIENCE CENTER

projects that we took on over 18 months. Forty-six of them required 151 data-analysis tasks. No project was identical, and we were surprised at how common one-off requests were (see 'Routinely unique'). There were a few routine procedures that many people wanted, such as finding genes expressed in a disease. But 79% of techniques applied to fewer than 20% of the projects. In other words, most researchers came to the bioinformatics core seeking customized analysis, not a standardized package.

This experience is not unique to our centre. "There are no 'one-size-fits-all' analyses," says Ian Korf, interim director of the Bioinformatics Core at the University of California, Davis.

Another issue is that projects usually become more complex as they go along. Often, an analysis addresses only part of a question and requires follow-up work. For example, when we unexpectedly found a weak correlation between a protein receptor and a signalling response, getting a more-robust answer required refining our analysis to accommodate distinct signalling pathways orchestrated by the protein. An expansion typically doubled the time spent on a project; in one case it increased the time fivefold.

Service cores such as ours are doing research. Our centre is more a partner than a routine service provider — the input affects the success of the project and shapes broader questions. As Korf puts it, "Our service isn't just doing. It's also part of the thinking."

PURE OR APPLIED

In our service core, surprisingly, only 5% of the time was spent on pure bioinformatics — that is, developing new algorithms that merit their own publications. Although we put together tools and analysis pipelines in creative ways, nearly all of our time was spent on applied bioinformatics.

The skill set required for each is different. Simon Andrews, head of bioinformatics at the Babraham Institute in Cambridge, UK, says that the most important factor for applied projects is interacting with biologists, "to know what goes on in the lab, and know what experiments are easy or hard."

Yet applied bioinformaticians do not lead their own research projects, so they do not slot cleanly into an independent faculty position, despite their PhD-level training. Service-core jobs are not seen to be as prestigious as faculty posts and offer fewer opportunities for advancement and leadership. Cores are not currently well placed to recruit researchers who have the necessary drive and expertise. Michael Rebhan, who leads a bioinformatics group at the Novartis Institutes for Biomedical Research in Basel, Switzerland, recalls his trouble setting up a helpdesk-type service. "It didn't work out and was difficult to attract really good scientists."

At the Bioinformatics Shared Resource at the University of Utah, co-director David Nix has found a solution: a separate career ladder for staff in his core, with titles and salary bands attached. Still, the potential for advancement is limited. The resource relies on people being determined to help others at the expense of professional prestige.

I have experienced these tensions myself. Although I do methodological research in my own lab, I spend increasingly more time on applying and adapting algorithms in the service core. This creates a richer research environment that boosts the institution's prestige. However, I am mainly evaluated on the success of my individual research programme. There is a need at my institute to engage in collaborative work, but this output is not captured by the standard metrics of achievement.

The research system does not recognize bioinformaticians for doing what the scientific community needs most. "People realize the importance, but currently there are no real solutions," says Xiaole Liu, a bioinformatician at the Dana-Farber Cancer Institute in Boston, Massachusetts, and at Tongji University in Shanghai, China. This is why it can take more than six months to fill positions at a core, why many of biology's brightest are leaving science for technology companies, and why conventional biologists wait nine months to get help to dissect their data.

GAP FILLING

Relatively little effort is needed to help train, recognize, reward and retain applied bioinformaticians. Physics, with its long history of massive, highly integrative projects has begun to sort out career paths, authorship norms and due credit for those in purely collaborative roles — most notably in beam-line service provision. The main barriers to doing so in bioinformatics are inertia and a lack of imagination. Organizations such as the International Society for Computational Biology and national funding bodies should take up the cause. Biologists should be vocal about the value their core collaborators bring.

In research institutes, there should be a formal career track optimized for applied bioinformatics. Like in tenure-track positions, promotions should be based on a body of work. A successful bioinformatician is one who contributes to successful collaborations, so evaluation schemes must recognize that applied bioinformaticians may never be corresponding or lead authors, but they are still accomplished scientists. Evaluation criteria should also consider whether a bioinformatician is adaptable, as shown by projects across diverse authors and specialities.

Granting agencies should develop principles to evaluate collaborative staff. When researchers apply for grants to cover our services, they want to name core-centre personnel who have published in the area of analysis. Reviewers should instead judge bioinformatics collaborators by their prior success in working with biologists. Naming these people on grant applications would give them deeper motivation for projects. Finally, there need to be more opportunities for biologists themselves to learn bioinformatics skills[2].

These steps should increase the pool of talented people who see applied bioinformatics as a promising career. Otherwise, research will stall. ■

Jeffrey Chang *is co-director of the Bioinformatics Service Center and assistant professor of integrative biology and pharmacology at the University of Texas Health Science Center at Houston, Texas, USA.*
e-mail: jeffrey.t.chang@uth.tmc.edu

1. Lewitter, F., Rebhan, M., Richter, B. & Sexton, D. *PLoS Comput. Biol.* **5,** e1000372 (2009).
2. Lewitter, F. & Rebhan, M. *PLoS Comput. Biol.* **5,** e1000368 (2009).

Caterina Sforza defended cities and hoarded alchemical secrets in the fifteenth century.

DE AGOSTINI PICTURE LIBRARY/AKG-IMAGES

HISTORY

Women at the edge of science

Jennifer Rampling assesses a study of the Renaissance women who dared to tackle alchemy's secrets.

Renaissance Italy had no female Galileo, but it did have Margherita Sarrocchi, a poet and avid student of mathematics, hailed by her contemporaries as a "marvel of the female sex". Seven of her letters to Galileo survive, along with one of his replies — relics of a correspondence in which Sarrocchi sought Galileo's opinion on her epic poem *Scanderbeide*, in return for her comments on his scientific work. We can only guess what form those comments might have taken. Sarrocchi's poem survives, but although her interest in geometry, astronomy and natural philosophy is widely attested, none of her work in these fields remains. The gap raises a question that resonates well beyond early modern Italy: how can we recover women's scientific knowledge, when even such an exceptional figure as Sarrocchi left no book for us to read?

Daughters of Alchemy: Women and Scientific Culture in Early Modern Italy
MEREDITH K. RAY
Harvard Univ. Press: 2015.

In *Daughters of Alchemy*, professor of Italian Meredith Ray seeks an answer in unexpected places: in women's poetry, letters, recipe collections, literary debates and alchemy. In early modern Europe, alchemy was pursued as both a branch of natural philosophy and a body of craft knowledge. This practical, vernacular dimension made it accessible to sixteenth- and seventeenth-century women at all levels of society. It is a world opened up in recent years by scholarship including Tara Nummedal's work on the sixteenth-century German alchemist Anna Zieglerin, and Alisha Rankin's *Panaceia's Daughters* (University of Chicago Press, 2013), on medical practice among German noblewomen. In it, women won alchemical patronage, tested and traded recipes, and adopted alchemical techniques such as distillation to prepare medicines, perfumes and transmuting elixirs.

Daughters of Alchemy brings together a series of case studies, beginning with the remarkable figure of Caterina Sforza, a noblewoman whose power bases were the northern Italian cities of Imola and Forlì. Sforza achieved lasting fame, and a mention in Niccolò Machiavelli's 1532 *The Prince*, for her defence of Forlì's citadel against the army of Cesare Borgia. As Ray shows, she also practised alchemy, using her knowledge of "secrets" as social capital. Sforza's *Experiments* was not intended for print, but was passed in manuscript to her son as a family heirloom. Later, the publishers of printed "books of secrets" sought to evoke the same sense of privileged, empirical knowledge. Indeed, the only female-authored book on nature from the time to become a best-seller, the 1561 *Secrets of Signora Isabella Cortese*, seems to have been written by a man seeking to exploit the rising market for women's secrets; Cortese is, after all, an anagram of *secreto*.

Women's associations with alchemy and secrecy, Ray argues, helped to carve out space in a wider literary arena. We encounter the Venetian poets Moderata Fonte and Lucrezia Marinella, who portrayed powerful female magi in their epic verse, and discussed science within an ongoing debate about the place of women in society. However, although Ray's basic claim is sound, she sometimes overstates the alchemical content of these works. When a character

in Fonte's dialogue *The Worth of Women* (1600) describes the interchangeability of the elements, she invokes a basic tenet of Aristotelian natural philosophy — not, as Ray suggests, a process "essentially alchemical in nature". Nor is it necessary to implicate the heterodox medical reformer Paracelsus in the commonplace image of the human body as a "little world". Fonte does innovate by proposing science as an area in which women can excel, but her actual views on natural philosophy seem to be mainstream.

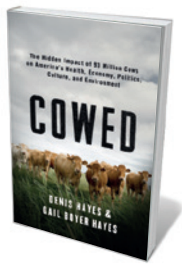
Fonte's orthodoxy contrasts with the attitude of another exceptional woman: Camilla Erculiani, an apothecary whose *Letters on Natural Philosophy* (1584) combined original scientific views (including a natural explanation for the biblical flood) with a staunch defence of women's intellectual capacity. Both author and practitioner, Erculiani sent her book to Poland for publication, armed with a dedication to the Polish queen, Anna Jagiellon. Yet even Krakow was not far enough away for her to escape the attention of the Inquisition. Humiliatingly, at her trial one of Erculiani's supporters used her sex to excuse her unorthodox reading of scripture — arguing that, as a woman, she could scarcely be held to know what she was talking about.

Female authors struggled to establish themselves as natural philosophers. Yet, Ray shows us, they could still make their voices heard as women, whether marketing women's secrets or defending their sex from slander. And they were effective. When Giuseppe Passi launched a misogynistic attack, *The Defects of Women* (1599), Marinella swiftly responded with her own book, *The Nobility and Excellence of Women and the Defects and Vice of Men* (1600). Among his jibes, Passi claimed that women could not 'do' science. Marinella out-argued him point by point. Eventually Passi backed off, retracting his more extreme views. Their exchange attracted an audience, and the books were reprinted.

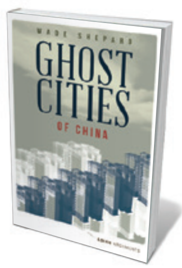
We can learn from this exchange. Marinella's treatise did not extend the boundaries of scientific knowledge, but that was not its aim. Such defences were important because they created an arena for women's voices. By writing about science, Renaissance women argued that they were qualified to write about science. Although they could not emulate Galileo in obtaining university posts, or seek roles as court philosophers, they could and did contribute to Renaissance scientific culture in other ways: as experimenters, readers, commentators, correspondents and critics. In sixteenth-century Europe, as in Ray's timely book, alchemy offered one way into a much larger conversation. ■

Jennifer Rampling *is assistant professor of history at Princeton University in New Jersey, where she teaches the history of alchemy.*
e-mail: rampling@princeton.edu

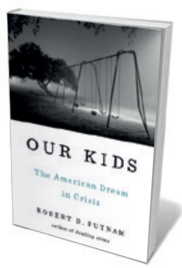
Books in brief



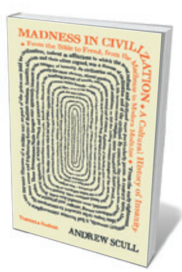
Cowed: The Hidden Impact of 93 Million Cows on America's Health, Economy, Politics, Culture, and Environment
Denis Hayes and Gail Boyer Hayes W. W. NORTON *(2015)*
Scattered among the 319 million US citizens are 93 million cows, supplying milk, beef and raw materials for substances from paint to toothpaste. But at a price: ruined soils, lagoons of excrement and significant greenhouse-gas emissions. So argue environmentalist Denis Hayes and environmental lawyer Gail Boyer Hayes in this richly researched overview. Marshalling numerous case studies, they show how humanity could shift from industrial farming to scaled-down, scientifically backed, sustainable animal husbandry.



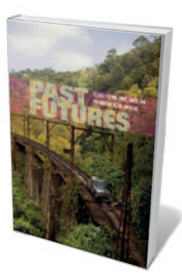
Ghost Cities of China: The Story of Cities without People in the World's Most Populated Country
Wade Shepard ZED *(2015)*
In 1949, China boasted 69 cities; now there are 657. Staggering in scale and set to churn on for 20 years, this experiment in urbanization is leaving a forlorn legacy: ghost towns that have yet to see an inhabitant. In this succinct study of a country bulldozed to make way for generic conurbations, *China Chronicle* editor Wade Shepard dispenses the facts with chilling clarity. As he examines mountains literally moved, relocation on a gargantuan scale and the duplication of Hallstatt, Austria, in Guangdong province, a stunned awe sets in.



Our Kids: The American Dream in Crisis
Robert D. Putnam SIMON AND SCHUSTER *(2015)*
Political scientist Robert Putnam's *Bowling Alone* (Simon and Schuster, 2000) exposed the increasing fragmentation of US communities. Now Putnam takes on the erosion of social mobility — once the keystone of the American dream. Meshing quantitative data and interviews with young people from the Deep South to the Rust Belt, he explores the class gap and finds that the vicious cycles of economic poverty often lead to political disengagement and lack of access to knowledge. His solutions, such as child tax benefits, inspire — but could founder without 1960s-style reformist zeal.



Madness in Civilization: A Cultural History of Insanity, from the Bible to Freud, from the Madhouse to Modern Medicine
Andrew Scull THAMES & HUDSON *(2015)*
In this ambitious chronicle of mental illness over two millennia, historian of psychiatry Andrew Scull ranges over the jumbled landscape of "Unreason" with crisp authority. His central argument is that the social and cultural contexts of extreme emotional states "dwarf any single set of meanings and practices". Insights jostle with horrors as Scull documents attempts to explain, contain and treat madness, from brutal asylums and lobotomies to the nuanced realization that mental illness has dual roots in society and biology.



Past Futures: Science Fiction, Space Travel, and Postwar Art of the Americas
Edited by Sarah J. Montross BOWDOIN/THE MIT PRESS *(2015)*
The science-fiction boom, cold war and space race of the mid-twentieth century set off a scientific and cultural explosion. Artists across the Americas discovered an alien splendour in the atomic age. This gripping volume showcases curator Sarah Montross's exhibition at Bowdoin College Museum of Art in Brunswick, Maine: from the cataclysmic (Rufino Tamayo's 1954 *Cosmic Terror*) to the rhythmic (Emilio Renart's 1965 *Drawing No. 13*), it is a revelation. **Barbara Kiser**

SCIENTIFIC INSTRUMENTATION

The aided eye

Philip Ball examines two studies on how optical instruments taught science to see.

In the seventeenth century, scientists learnt how to see, discovering the astronomically large and the invisibly small. Both the telescope and the microscope had been invented, independently, by the first decades of the century, and Europe's intelligentsia were astonished, amused and unnerved by what was revealed.

In *Eye of the Beholder*, historian Laura Snyder describes the insights derived from the microscope by Dutch cloth merchant Antoni van Leeuwenhoek, who, using self-made microscopes with a resolution as fine as one micrometre, found teeming life in drops of rainwater. In *Galileo's Telescope*, historians of science Massimo Bucciantini, Michele Camerota and Franco Giudice offer a new account of how Galileo Galilei introduced the world to the telescope's power to unravel the heavens. They track the genesis and influence of Galileo's 1610 booklet *Sidereus nuncius* (*Starry messenger*). Both of these detailed studies show how sensational it was to discover worlds not perceivable to the naked eye.

The Astronomer (1668) may depict Antoni van Leeuwenhoek.

AKG-IMAGES

Snyder also explores the parallels between the interests of Leeuwenhoek and those of the artist Johannes Vermeer. Both men of Delft, they put lenses to work for different purposes — Leeuwenhoek to satisfy an insatiable curiosity, Vermeer to extend his ability to perceive and record the world, for example with a camera obscura. Did they share knowledge as acquaintances, even friends? Leeuwenhoek was executor of Vermeer's estate; although this may have been the civic duty of an eminent merchant, Snyder points out that the few other times Leeuwenhoek took such a role, he had links with the deceased.

It has been suggested that the scholar in two of Vermeer's paintings from the 1660s — with map and dividers in *The Geographer* and with a globe in *The Astronomer* — was Leeuwenhoek. Known portraits of him are from a later date, so the resemblance is hard to judge, and Snyder is unable to settle the matter. She can only speculate about whether Vermeer inspired Leeuwenhoek to use lenses for more than assessing the value of fabrics.

Nonetheless, Snyder beautifully evokes the ambience of late-seventeenth-century Delft, "where an obliging butcher would sell Leeuwenhoek cows' eyes, the testicles of hares, and other required specimens". She is revelatory about Vermeer's aims and methods, helping to explain what is so mesmeric about his work. That sent me scampering to examine the "specular highlights" of the bread in *The Milkmaid*, produced with refined layering of pigments. "What Vermeer was painting was the way the eye actually sees, not the way the mind thinks it sees," she writes.

That distinction underlies both books, and encapsulates the contention over what observation meant. It is sometimes assumed that the introduction of new instruments was unproblematic to all but the bigoted and ignorant. In fact, the first telescopes and microscopes failed to reveal much. You needed to 'get your eye in' before you could interpret what you saw.

Diarist Samuel Pepys admitted "great difficulty before we could [see] any thing" in the microscope he bought after reading natural philosopher Henry Power's description in 1664. Even Robert Hooke, asked by the Royal Society to verify Leeuwenhoek's claims, found it difficult to use the merchant's favoured single-lens microscopes, which give better magnification than the compound microscopes that Hooke had used for his *Micrographia* (1665). And Galileo wondered, when Saturn shifted and its rings became less clear, whether his instrument was deceiving him. There was debate over whether such devices could be trusted. Optics had a disreputable association with magic: the Neapolitan Giambattista della Porta, who had perpetuated the link in his 1558 book *Natural Magic*, initially dismissed Galileo's claims (while taking credit for inventing the telescope): "I have seen the secret use of the eyeglass and it's a load of balls."

Both Leeuwenhoek and Galileo were secretive and possessive about their devices. Galileo mastered lens-grinding to improve on Dutch instruments (he relied on verbal descriptions). But he was determined to keep his own telescopes from rivals: entreaties from Johannes Kepler, with whom he was on good terms, went unheeded. And there was scant understanding of how the instruments worked. Galileo said airily in *Sidereus nuncius* that he had perfected the device "on the basis of the science of refraction", yet it was Kepler who first gave a fair account of the principles, in his 1611 *Dioptrice*.

Galileo's Telescope restricts itself to the period when Galileo became famous for descriptions of the surface of the Moon, the Milky Way "powdered with stars" (as writer John Milton described it), and the moons of Jupiter — which he called the Medicean stars to flatter Cosimo II de' Medici, his patron. The challenge this complexity posed to traditional cosmology foreshadowed theological storms to come. As the authors put it: "If the sky was subject to generation and corruption, could it continue to be the home of angels and saints?"

Galileo's Telescope is not a light read: more context amid the richly researched detail would have helped, and there is little on how Galileo's personality shaped his debates. But both books project a sense of how new ways of seeing, far from merely providing new tools, were — and are — complicated extensions of the way we understand our experience. ■

Eye of the Beholder: Johannes Vermeer, Antoni van Leeuwenhoek, and the Reinvention of Seeing
LAURA J. SNYDER
W. W. Norton: 2015.

Galileo's Telescope: A European Story
MASSIMO BUCCIANTINI, MICHELE CAMEROTA & FRANCO GIUDICE, TRANSLATED BY CATHERINE BOLTON
Harvard Univ. Press: 2015.

Philip Ball *is a writer based in London. His latest book is* Invisible.
e-mail: p.ball@btinternet.com

Correspondence

Academies review insecticide harm

The European Academies Science Advisory Council (EASAC) will next week release its report 'Ecosystem services, agriculture and neonicotinoids', which scrutinizes the scientific evidence for harmful effects by neonicotinoid insecticides. It concludes that widespread preventive use of neonicotinoids has adverse effects on non-target organisms that provide ecosystem services such as pollination and natural pest control.

The EASAC report goes beyond honeybees to include other valuable pollinators, such as bumblebees and solitary bees, and looks at ecosystem services that are crucial to sustainable agriculture. It is based on the findings of an international group of independent scientists, which I chaired, with expertise ranging from pollination biology through systems ecology to toxicology (see www.easac.eu).

The report points out that the preventive use of neonicotinoids is inconsistent with the principles of integrated pest management, as expressed in the European Union's (EU) sustainable pesticides directive. Such usage also constrains the potential for restoring farmland biodiversity under the EU agri-environment regulation. The group notes that neonicotinoids also have sublethal effects that need to be fully addressed in EU approval procedures.

The European Commission is due to review the effects of its 2013 restriction on the use of neonicotinoids on flowering crops. I believe that our report will help the European review to reassess the risk–benefit balance of neonicotinoid application. The wider risks to the environment and longer-term sustainability of agriculture must be considered alongside concerns that further restrictions could have short-term implications for the economy and for food security.

Peter Neumann *Institute of Bee Health, University of Bern, Switzerland.*
peter.neumann@vetsuisse.unibe.ch

Sanctions in Iran disrupt cancer care

Economic sanctions against Iran have not formally targeted health care or access to drugs, but they have indirectly led to serious problems for health services — notably for people with cancer.

Cancer is the third-highest cause of death in Iran, greater than in most other Middle Eastern countries. The Program of Action for Cancer Therapy, established by the International Atomic Energy Agency, evaluated the status of Iran's National Cancer Control Program (NCCP) in 2012. It concluded that the NCCP has substantial deficits in all aspects of care, including prevention, early detection, diagnosis, treatment, palliative care and monitoring technology (M. R. Rouhollahi *et al. Arch. Iran Med.* **17,** 222–231; 2014). There is also a serious shortage of cancer drugs in the nation.

Sanctions disrupt health services and basic nutrition through complications in transportation, reduced imports, and difficulty in transferring hard currencies. There can be no improvement in the status of the NCCP until they are lifted.

Shohreh Shahabi* *Danbury Hospital, Connecticut; and Columbia University, USA.*
shohreh.shahabi@wchn.org
**On behalf of 5 correspondents (see go.nature.com/gsdvk3 for full list).*

No case for Japan to kill minke whales

On 12 April, an expert panel convened by the Scientific Committee of the International Whaling Commission (IWC) will deliver its review of Japan's latest proposal to kill Antarctic minke whales for research. Along with other members of the panel, I find that the Japanese government's scientific case does not justify lethal capture of any whales.

Japan's proposal to kill 333 whales every year until 2026 was submitted after the International Court of Justice ruled in March 2014 that the Japanese Whale Research Program under Special Permit in the Antarctic II (JARPA II) was not for purposes of scientific research. The court ordered Japan "to revoke any extant authorization, permit or licence to kill, take or treat whales in relation to JARPA II".

It is essential, therefore, that the review outcome is decisive and depends only on the quality of the scientific case for whaling made by Japan. This is implicit in the commission's new terms of reference (see IWC *J. Cetacean Res. Manage.* **16,** S1–S5; 2015 and go.nature.com/j4dbeg).

Under these new terms, review panels are not required to make suggestions on how to improve proposals — but this historical practice still persists. I believe strongly that it must cease. It should not be assumed that whaling is inevitable (see P. J. Clapham *Mar. Policy* **51,** 238–241; 2015). Neither should panel members be made collaborators in an iterative approach towards approving whaling proposals. Otherwise, there is a risk that IWC practice as currently in effect could subvert the norm of independent review.

Andrew S. Brierley *University of St Andrews, UK.*
asb4@st-and.ac.uk

One scientist's data as another's noise

Reviewer Charles Seife considers my book *Big Data, Little Data, No Data: Scholarship in the Networked World* alongside two popular books on big data in commerce: Steve Lohr's *Data-ism* and Bruce Schneier's *Data and Goliath* (see *Nature* **518,** 480–481; 2015). Yet the review does not make clear that mine is aimed at a very different audience.

I wrote the book for scientific researchers, scholars, librarians, publishers, policy-makers and other stakeholders, for whom the subtle uses of data as evidence in research are being swamped in the hype about big data. With case studies exploring how the idea of data varies in and between domains, I show how one researcher's data can be someone else's noise. Therein lies the rub.

Because so much about research data is open to personal interpretation, the information can be difficult to describe, represent and manage — and to share or reuse. The failure to understand these complexities leads to misguided policies for data management and to a lack of investment in both the workforce and the infrastructure for data curation. Ultimately, it can mean that no data survive for research.

Christine Borgman *University of California, Los Angeles, USA.*
cborgman@g.ucla.edu

Another two-body trick before Dawn

You claim that NASA's Dawn probe is the first to have orbited two extraterrestrial bodies (*Nature* **519,** 134; 2015), having arrived at the dwarf planet Ceres on 6 March after orbiting the asteroid Vesta in 2011–12 (see also *Nature* http://doi.org/287; 2015). But the Galileo mission to Jupiter had already accomplished the two-body trick.

Leaving aside a Copernican view that every space probe orbits the Sun even before launch, a gravity-assist trajectory passing Venus (once) and Earth (twice) put the Galileo craft in orbit around the Sun for more than three years. After heading to Jupiter, Galileo orbited it for almost eight years before plunging into the planet in 2003 to protect Jupiter's moon Europa from contamination.

John D. Rummel *East Carolina University, Greenville, North Carolina, USA.*
rummelj@ecu.edu

CARDIOLOGY

Race for healthy hearts

Transplantation experiments in mice reveal that the increased risk of congenital heart disease in the pups of older mothers is not conferred by ageing eggs, but by the mothers' age, and can be mitigated by exercise. SEE LETTER P.230

MARC-PHILLIP HITZ
& GREGOR ANDELFINGER

Congenital heart disease is the most common developmental malformation in humans, and a leading cause of death in infancy. About 1 in 100 children have minor congenital heart disease, and 1 in 1,000 children will need heart surgery. Epidemiological studies[1,2] indicate that there are many risk factors for congenital heart disease: infections, genetics, environmental factors such as toxins or maternal diabetes, and higher maternal age, to name just a few. In this issue, Schulkey *et al.*[3] (page 230) question whether, in terms of age, it is that of the egg or of the mother herself that affects the risk of congenital heart disease in offspring of older mothers. Their answer is clear — the risk lies, surprisingly, in the age of the mother.

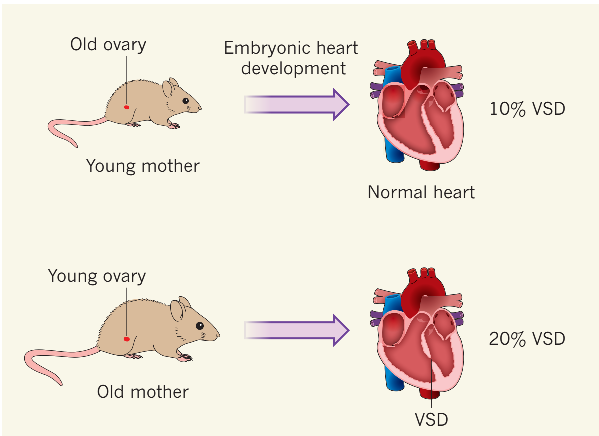


Figure 1 | The maternal influence. Schulkey *et al.*[3] transplanted old ovaries into young female mice, and vice versa. They found that 90% of young mice gave birth to pups that had normal hearts, with only 10% developing a ventricular septal defect (VSD). By contrast, 20% of pups from old mothers developed VSD, indicating that the age of the mother is a major determinant of the incidence of this congenital defect.

The heart is the first organ to form in the vertebrate embryo, following a precisely orchestrated and evolutionarily conserved developmental process[4]. In mice, septation of the heart into a four-chambered structure consisting of two atria and two ventricles begins with the formation of the interatrial and interventricular septa[5]. A master regulator of this process is Nkx2-5, a cardiac homeobox transcription factor. The gene that encodes this protein is frequently mutated in humans with septation defects[6].

Schulkey and colleagues took advantage of mice that have only one functional copy of the *Nkx2-5* gene, instead of the usual two, and that as a result are more likely to develop ventricular septal defects (VSDs)[7]. The authors performed reciprocal ovarian transplants, in which ovaries from young mice were transplanted to older mice and vice versa, and convincingly showed that the risk of VSD increases with the advancing age of the mother, but not of the ovaries — and therefore, not of the egg (Fig. 1). Next, to rule out the main environmental factors implicated in epidemiological studies of congenital heart defects (maternal diet, body mass and glucose intolerance), pregnant mice were fed a high-fat diet. A higher body weight and altered glucose metabolism did not change VSD incidence in young mothers, and did so only marginally in older mothers.

Where, then, does the VSD risk originate, and can it be mitigated? To examine whether genetic factors are also involved in risk, the authors bred their mice with those from other strains with different genetic make-ups (genetic backgrounds). Changes in genetic background modified the risk of maternal-age-associated VSD, thus providing evidence for a complex gene–environment interaction. Finally, and importantly, Schulkey *et al.* demonstrated that VSD risk can be considerably decreased by placing mice in running wheels. Small amounts of exercise were insufficient, but beyond a threshold amount, exercise had a protective effect.

This study is relevant for several reasons. Because the authors' mouse model recapitulates the complex nature of human congenital heart disease, it makes individual contributors of disease risk quantifiable. The reciprocal transplant protocol will be instrumental in studying whether maternal-age-associated risk is also a factor in mice harbouring other genetic mutations, and whether the risk is identical in subsequent pregnancies. Schulkey and colleagues' model will also enable us to identify the genetic and environmental modifiers associated with risk, and to devise intervention strategies to make human pregnancies safer.

At present, we can only speculate about the nature of the factors that increase the maternal-age-related risk of VSDs. On the basis of what is known about heart development, and of constraints imposed by the design of the current study, we can assume that the elusive factor or factors must be embryo-independent, maternally based, influenced by the maternal genome and affected by exercise with a threshold effect. Two possibilities spring to mind: metabolic and epigenetic factors.

Epigenetic factors alter gene expression, without changing the underlying DNA sequence, in a permanent manner that can be passed down through generations of cells and, in sperm and eggs, through generations of organisms. Such transgenerational epigenetic inheritance has been described in the pups of mice with reduced activity of the gene *Mtrr* — developmental heart defects arise not only in mutant offspring, but also in up to four generations of wild-type descendants[8]. Transfer of embryos habouring the *Mtrr* mutation into wild-type mothers revealed two separable traits: growth defects caused by the uterine environment, and congenital abnormalities passed on through epigenetic inheritance. The underlying molecular processes are unclear. However, the defects brought about by *Mtrr* mutation can be mitigated by supplementing the maternal diet with folate, and evidence[9] suggests that folate supplementation might work by altering the patterns of an epigenetic modification called methylation. Whether the influence of age and exercise on epigenetic effects is independent of, or under the control of, folate levels could be tested experimentally using Schulkey and colleagues' model. Controlled interventional studies[10,11] show that exercise can change the methylation levels of several genes and pathways, including those involved in metabolism, signalling and transcription.

Several questions remain to be answered. First, what is the precise role of Nkx2-5 in the maternal placenta and in those tissues derived

from the fertilized egg that do not form the embryo, such as the yolk? A step towards answering this question would be to determine whether the outcome remains the same when embryos instead of ovaries are transplanted — although, admittedly, the sheer magnitude of such an undertaking seems prohibitive. Second, what about paternal mutations, which were not analysed in this study? Third, what is the role of other types of age-related genetic factor, such as mutations resulting in placental growth disturbances? A wide range of factors, including age-related decline in metabolic or placental functions, could conceivably contribute to such anomalies. But on the basis of both Schulkey and co-workers' model and another recent study[12], it seems clear that these factors must be under genetic control.

Future studies will need to be large, and to focus on multigenerational families, if they are to provide sufficient statistical evidence for genetic risk. Studies similar to Schulkey and colleagues' should be designed to better quantify epigenetic risk factors. Subsequently, the fraction of risk attributable to different factors needs to be defined, in particular those amenable to interventional strategies, such as the use of folate supplements. In light of the increasing average maternal age at childbirth, and the advent of egg freezing as an employee benefit in some countries, the authors' study provides a timely opportunity to improve risk stratification and management of the stages immediately before and after conception, to prevent birth defects. ■

Marc-Phillip Hitz *is at the Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK, and at the University Hospital Schleswig-Holstein and the Christian-Albrechts University, Kiel, Germany.*
Gregor Andelfinger *is in the Department of Pediatrics, CHU Sainte Justine Research Center, Université de Montréal, Montreal H3T 1C5, Canada.*
e-mail: gregor.andelfinger@recherche-ste-justine.qc.ca

1. Miller, A., Riehle-Colarusso, T., Siffel, C., Frías, J. L. & Correa, A. *Am. J. Med. Genet. A* **155,** 2137–2145 (2011).
2. Ferencz, C., Rubin, J. D., Loffredo, C. A. & Magee, C. A. (eds) *The Epidemiology of Congenital Heart Disease: The Baltimore-Washington Infant Study 1981–1989* (Futura, 1993).
3. Schulkey, C. E. *et al. Nature* **520,** 230–233 (2015).
4. Lescroart, F. *et al. Nature Cell Biol.* **16,** 829–840 (2014).
5. De La Cruz, M. V., Giménez-Ribotta, M., Saravalli, O. & Cayré, R. *Am. J. Anat.* **166,** 63–72 (1983).
6. Kasahara, H., Seidman, C. E. & Izumo, S. *J. Clin. Invest.* **106,** 299–308 (2000).
7. Winston, J. B. *et al. Circ. Cardiovasc. Genet.* **5,** 293–300 (2012).
8. Padmanabhan, N. *et al. Cell* **155,** 81–93 (2013).
9. Ionescu-Ittu, R. *et al. Br. Med. J.* **338,** b1673 (2009).
10. Nitert, M. D. *et al. Diabetes* **61,** 3322–3332 (2012).
11. Rönn, T. *et al. PLoS Genet.* **9,** e1003572 (2013).
12. Mitchell, L. E. *et al. Hum. Mol. Genet.* **24,** 265–273 (2015).

This article was published online on 1 April 2015.

PLANETARY SCIENCE

Prebiotic chemistry on the rocks

Organic compounds called nitriles have been detected in material surrounding a young star. The finding hints at a vast reservoir of ice and volatile species that can seed the surfaces of young rocky planets or moons. SEE LETTER P.198

GEOFFREY A. BLAKE & EDWIN A. BERGIN

The recipe for creating a habitable planet such as Earth contains several essential ingredients. Once the key components — a silicate mantle and a metallic, iron-rich core — have been built, there must be sufficient liquid water and appropriate forms of carbon and nitrogen available near the surface, along with a sprinkling of sulfur and phosphorus. Water and the organic carriers of the necessary elements (C, N, S, P) are known as volatiles because temperatures must be very low (below about 150 kelvin) for them to be frozen into the pebbles that are the seeds of rocky worlds. On page 198 of this issue, Öberg *et al.*[1] report the discovery of spectral emission lines from gaseous molecules of C–N-containing organic species in potentially planet-forming environments using the Atacama Large Millimeter/submillimeter Array (ALMA). The observed lines trace the surface of a vast reservoir of icy bodies that can deliver volatile organics to the surfaces of young rocky planets or to moons circling gas-giant planets at distances from the central star at which liquid water is stable.

Stars are born in giant clouds of gas and dust, such as the famous Orion nebula, which is part of the 'sword' in the constellation of Orion. These clouds host a complex web of chemistry involving hundreds of molecular species[2], most of them organic. The particular environment under study here is a 'protoplanetary disk' of gas, dust and ice surrounding a young (that is, a few million years old) star called MWC 480 in the constellation of Taurus. Once a young star is nearly fully assembled, these disks rotate in Keplerian orbital motion, and are the birthplaces of planetary systems. The physical conditions in disks vary greatly, with hot and dense regions of gas and dust near to the star and much colder gas, dust and eventually ice at greater distances from it[3].

To build planets, material inherited from

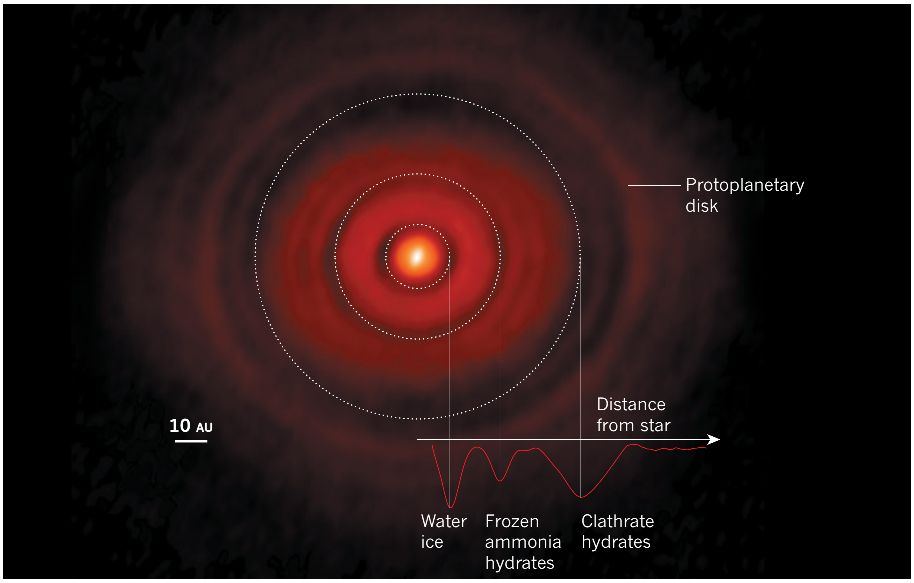


Figure 1 | HL Tau disk. The image shows a face-on millimetre-wave view of the protoplanetary disk around the young star HL Tau (located at the bright central blob but not detected in this image). The dotted circles represent the locations of emission dips (dark regions) where water ice, frozen ammonia hydrates and clathrate hydrates are expected to condense. Combining such images, which can directly probe the disk's dust and ice content, with spectral-line observations such as those presented by Öberg *et al.*[1] for the disk around the young star MWC 480, which selectively probe the gas and ice reservoirs of the disk, will provide insights into the evolution of water and organic chemistry during planet formation. 1 AU is the distance from Earth to the Sun. (Image created by K. Zhang in G. A. Blake's research group, from public-domain commissioning and science verification data from the ALMA observatory[8].)

the natal cloud travels radially and vertically through diverse conditions in the disk. Small grains of ice and dust grow larger and settle vertically, eventually forming objects such as asteroids and comets in the densest, mid-plane region of the disk — a zone that is extremely difficult to study directly because it is largely obscured by the overlying dust, ice and gas. Near the surface of the inner part of the disk, close to the central star, conditions are warm enough for spectral emission from simple volatiles to be detected at infrared wavelengths[4]. However, such data do not reveal whether more-complex species are present, nor what chemistry prevails in the outer regions of the disk, where icy, kilometre-sized or larger bodies known as planetesimals should form and eventually coalesce into even larger bodies. Enter ALMA and the pioneering observations of Öberg and colleagues.

In their study, Öberg *et al.* have detected spectral emission lines associated with the rotational states of three C–N-containing species — hydrogen cyanide (HCN), acetonitrile (CH_3CN) and cyanoacetylene (HC_3N) — in the disk surrounding MWC 480 at distances between 30 and 100 astronomical units from the central star (1 AU is the distance from Earth to the Sun). At these distant and cold radii, such nitrile compounds should be locked into icy dust grains or planetesimals.

MWC 480 has roughly twice the mass of the Sun and is brighter. The distances probed compare well with those over which comets were assembled in our own Solar System. The new ALMA data sample the atmosphere of the outer parts of the protoplanetary disk, where only highly volatile species can remain in the gas phase, and where icy grains, lofted high into the disk atmosphere from the cold mid-plane, can interact with ultraviolet photons from the young star, driving species trapped on the grain surfaces into the gas phase. The nitriles fall into the latter category, and are central to prebiotic chemistry, because they probably represent precursors of more-complex species, such as amino acids.

By modelling the observed emission, Öberg *et al.* concluded that the organic nitriles must be abundant in the disk ices, even more so than is currently observed in comets. It has been known for some time that primitive Solar System bodies inherit starting materials from the earlier stages of the star-formation process. But this work, along with other recent studies (see, for example, ref. 5), demonstrates that protoplanetary disks are active engines of chemical synthesis, and that such environments are vital for building chemical complexity long before a planetary surface is created. The potentially prebiotic chemistry traced by asteroids and comets in the Solar System is therefore replicated, at least in part, in other young planetary systems — suggesting that planets are supplied with these life-bearing elements as they are born.

We stand to learn a great deal about the early steps of planet formation and the chemistry of volatiles in the coming decade, particularly as ALMA ramps up to its full capabilities. A spectacular early hint of what is to come is provided by the ALMA millimetre-wave imaging data released earlier this year for HL Tau (Fig. 1), a young star with substantial surrounding cloud material. It would be surprising for planet formation to be at an advanced stage in HL Tau's disk. The high gas-accretion rate in such young stars can move the snowline radii, beyond which certain molecules condense, out to large distances from the star[6], making the snowlines potentially easier to study in such objects. One possible explanation for the emission dips observed by ALMA (Fig. 1) is that they correspond nicely to the expected locations for the condensation radii of water ice, frozen ammonia (and hydrogen sulfide) hydrates and clathrate hydrates (which contain carbon dioxide, methane, carbon monoxide or nitrogen gas) — locations at which rapid pebble growth has been predicted to occur[7]. As the pebbles' diameters grow to be substantially larger than the observing wavelength, the dust and ice emission would drop.

By combining infrared spectra with high-resolution, millimetre-wave imaging data, and with spectral-line observations such as those described by Öberg and co-workers to probe the upper layers of the disk, it will be possible to determine the speciation of the most abundant volatile species, examine their distributions, and assess the likelihood of their delivery to nascent planetary surfaces. ■

Geoffrey A. Blake *is at the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.* **Edwin A. Bergin** *is in the Department of Astronomy, University of Michigan, Ann Arbor, Michigan 48109, USA.*
e-mails: gab@gps.caltech.edu; ebergin@umich.edu

1. Öberg, K. I. *et al. Nature* **520**, 198–201 (2015).
2. Herbst, E. & van Dishoeck, E. F. *Annu. Rev. Astron. Astrophys.* **47**, 427–480 (2009).
3. Henning, T. & Semenov, D. *Chem. Rev.* **113**, 9016–9042 (2013).
4. Carr, J. S. & Najita, J. R. *Science* **319**, 1504–1506 (2008).
5. Bergin, E., Cleeves, L. I., Crockett, N. & Blake, G. *Faraday Discuss.* **168**, 61–79 (2014).
6. Men'shchikov, A. B., Henning, T. & Fischer, O. *Astrophys. J.* **519**, 257–278 (1999).
7. Ros, K. & Johansen, A. *Astron. Astrophys.* **552**, A137 (2013).
8. ALMA Partnership. Preprint at http://arxiv.org/abs/1503.02649 (2015).

MOLECULAR BIOLOGY

RNA interference hangs by a thread

The Paf1 protein complex in fission yeast has been found to protect protein-coding genes from inhibition by RNA-mediated silencing of transcription, by stimulating the release of nascent transcripts from DNA. **SEE LETTER P.248**

MIKEL ZARATIEGUI

RNA interference (RNAi) is a mechanism of gene regulation that uses small RNAs called short interfering RNAs (siRNAs)to silence the expression of specific targets that have complementary nucleotide sequences. This can occur through post-transcriptional silencing, which degrades the target transcript, or through modification of chromatin (the complex of proteins, RNA and DNA in which DNA is packaged in the cell), which prevents transcription from initiating. RNAi always causes transcript degradation, but its effect on chromatin is bewilderingly inconsistent; only some targets in some organisms exhibit RNAi-mediated chromatin modifications, despite the evolutionary conservation of this mechanism. In this issue, Kowalik *et al.*[1] (page 248) report that the target transcript must remain associated with the site of transcription for RNAi to act on chromatin, providing a possible explanation for this variability.

Researchers can exploit RNAi to inhibit any RNA sequence of choice, simply by introducing a source of siRNA, such as a double-stranded RNA molecule that can be processed to siRNA by the nuclease enzyme Dicer. During transcription, siRNA can hybridize with a complementary nascent transcript and induce the deposition of silencing chromatin marks, leading to the formation of an inheritable repressive type of chromatin called heterochromatin. The inheritance of silencing chromatin modifications down generations of cells is an example of epigenetic memory.

By forming complexes with Argonaute effector proteins, siRNA can locate and silence target sequences even if they are located at distant sites, a process called silencing *in trans*. This phenomenon is crucial for repressing the

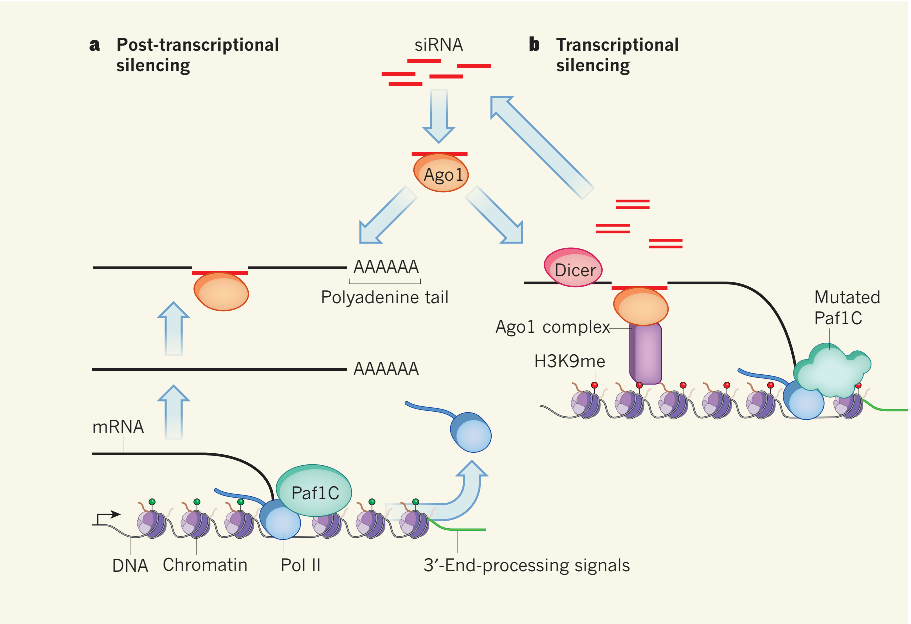


Figure 1 | Interfering with interference. Short interfering RNA (siRNA) forms complexes with the Argonaute 1 (Ago1) protein, enabling it to bind to and cleave complementary transcripts. **a**, In fission yeast, the Paf1 protein complex (Paf1C) promotes both transcript elongation by the enzyme RNA polymerase II (Pol II) and 3′-end processing, which consists of the release of Pol II, and cleavage and polyadenylation of messenger RNA, releasing the mRNA from the site of transcription. The released transcript is silenced post-transcriptionally by siRNA. **b**, Kowalik *et al.*[1] report that, if mutations compromise Paf1C activity, the uncleaved transcript is retained at the site of transcription. Ago1–siRNA forms complexes with other proteins, which together mediate the induction of chromatin-silencing modifications such as H3K9me, and the initiation of a self-reinforcing cycle of secondary siRNA production, in which mRNA is cleaved by siRNA, and the cleaved transcript is processed to siRNA by the nuclease enzyme Dicer.

transcription of dispersed repetitive genomic elements, such as transposons[2], which must be silenced to prevent them from 'jumping' around the genome and introducing harmful mutations. However, when researchers try to take advantage of this phenomenon to induce heterochromatin in mammalian protein-coding genes, the process seldom works. Instead, it produces erratic results that depend in perplexing ways on the target gene and on pre-existing chromatin modifications.

Similarly disappointing results have been reported for fission yeast (*Schizosaccharomyces pombe*). The yeast's siRNA precursors cannot induce heterochromatin formation in target genes[3], although certain non-protein-coding sequences at the centre of the yeast's chromosomes are robustly silenced by RNAi. Reasoning that suppressive mechanisms might protect protein-coding genes from RNAi, Kowalik and colleagues performed a genetic screen in *S. pombe*, looking for mutations that allowed silencing *in trans*. Strikingly, all the mutations they obtained arose in proteins of the RNA polymerase-associated factor 1 complex (Paf1C).

The authors report that several different siRNA precursors can reliably induce heterochromatin formation when Paf1C is mutated. The silenced DNA regions exhibit all the hallmarks of endogenous heterochromatin, including decreased transcription, generation of target-derived (secondary) siRNA molecules, and addition of a methyl group to the amino-acid residue lysine 9 of the chromatin protein histone 3 (a modification known as H3K9me). Kowalik *et al.* found that, once established, siRNA-induced heterochromatin can be inherited across generations of cells, even in the absence of the original siRNA source, as long as Paf1C is still mutated. This suggests that the cycle of secondary siRNA generation and H3K9me deposition that arises after heterochromatin initially forms can induce its own inheritance (Fig. 1). Silencing marks are re-established after each chromatin replication, preventing their dilution and satisfying even the strictest definitions of epigenetic inheritance.

How does Paf1C prevent RNAi from silencing chromatin? The authors tackled this question by genetically inhibiting proteins involved in each of the three main functions regulated by Paf1C: chromatin modification; elongation of messenger RNA during transcription; and processing of the 3′ end of mRNA as transcription terminates[4]. During 3′-end processing, transcripts are cleaved and a 'polyadenine tail' is added to stabilize the mRNA and permit normal translation. Interfering with chromatin modification or transcript elongation had no effect on heterochromatin formation. However, mutations in Ctf1 and Res2, two proteins required for cleavage and polyadenylation of mRNA, allowed *in trans* silencing to occur, albeit less efficiently than in the Paf1C mutants.

Conversely, Kowalik *et al.* demonstrate that bypassing 3′-end processing, by using a ribozyme molecule that cleaves off and releases the nascent transcript from DNA before termination, prevents *in trans* silencing even when Paf1C is mutated. Thus, it seems that the expeditious 3′-end processing of protein-coding genes protects them from RNAi by denying siRNA a docking platform — a nascent transcript still attached to the site of transcription. This process is probably necessary to prevent spurious heterochromatin formation by runaway RNAi activity, because the authors observe stochastic formation of heterochromatin in abnormal, protein-coding locations when Paf1C activity is compromised.

On a fundamental level, these results highlight the inextricable relationship between RNAi and mRNA processing. Both share the same targets, which are transcribed by the enzyme RNA polymerase II, and the two processes share much cross-regulation. It seems that the two mechanisms converge in 3′-end processing. Almost a decade ago, genetic screens searching for enhancers of post-transcriptional RNAi in plants yielded factors involved in 3′-end processing[5], a harbinger of the effect shown here for transcriptional silencing. Strikingly, it seems that Dicer can release some nascent transcripts[6,7], usurping the role of the normal cleavage and polyadenylation machinery in certain situations. Together with the known influence of chromatin state on 3′-end processing[8], a picture is emerging of a network of regulatory activities that enhance or prevent chromatin silencing in different targets.

But the most exciting implication is the potential to leverage this knowledge to induce targeted heterochromatin formation in different organisms, including humans. Mechanisms of RNAi-mediated chromatin modification might be universally conserved after all, with the variability observed in different targets and models stemming from differences in the efficiency of nascent transcript release at target sequences. In *S. pombe*, the transcripts naturally targeted for RNAi-mediated heterochromatin formation exhibit notoriously inefficient termination of transcription[9], which explains why their silencing is independent of Paf1C activity. Furthermore, a protein-coding reporter gene can be efficiently silenced *in trans* in this yeast if its 3′-end-processing signals are mutated[10]. Perhaps, then, we will be able to silence genes reliably and inheritably at will, by inhibiting transcription termination[11] and simultaneously using a specific RNAi trigger. In this way, RNAi would finally fulfil its promise and join other technologies, such as CRISPR/dCas9 systems, that enable editing of the epigenome. ■

Mikel Zaratiegui *is in the Department of Molecular Biology and Biochemistry, Rutgers, the State University of New Jersey, Piscataway, New Jersey 08854, USA.*
e-mail: zaratiegui@dls.rutgers.edu

1. Kowalik, K. M. *et al. Nature* **520,** 248–252 (2015).
2. Sienski, G., Dönertas, D. & Brennecke, J. *Cell* **151,** 964–980 (2012).
3. Sigova, A., Rhind, N. & Zamore, P. D. *Genes Dev.* **18,** 2359–2367 (2004).
4. Tomson, B. N. & Arndt, K. M. *Biochim. Biophys. Acta Gene Reg. Mech.* **1829,** 116–126 (2013).
5. Herr, A. J., Molnàr, A., Jones, A. & Baulcombe, D. C. *Proc Natl Acad. Sci. USA* **103,** 14994–15001 (2006).
6. Liu, F., Bakht, S. & Dean, C. *Science* **335,** 1621–1623 (2012).
7. Castel, S. E. *et al. Cell* **159,** 572–583 (2014).
8. Gullerova, M. & Proudfoot, N. J. *Cell* **132,** 983–995 (2008).
9. Zaratiegui, M. *et al. Nature* **479,** 135–138 (2011).
10. Yu, R., Jih, G., Iglesias, N. & Moazed, D. *Mol. Cell* **53,** 262–276 (2014).
11. Nemeroff, M. E., Barabino, S. M. L., Li, Y., Keller, W. & Krug, R. M. *Mol. Cell* **1,** 991–1000 (1998).

This article was published online on 25 March 2015.

MATERIALS SCIENCE

Unique wrinkles as identity tags

Spontaneously generated, random wrinkles of coatings on microscale particles have been found to be analogous to fingerprints — unique patterns with a wavy topography that can serve as unclonable tags for anti-counterfeiting purposes.

JIE YIN & MARY C. BOYCE

Gottfried Wilhelm Leibniz articulated his principle of the indiscernibility of identicals[1] with the words "No two identical leaves exist in the garden". Writing in *Advanced Materials*, Bae *et al.*[2] demonstrate the same principle for wrinkled microscale particles that have disordered 'labyrinth' surface patterns — no two identical microparticles form even when generated under seemingly identical conditions. The authors have used the unique and irreproducible characteristics of these wrinkles as "artificial microfingerprints" for identification and encryption applications.

Wrinkles are ubiquitous in our lives, from the wavy undulations of our skin, to the folds of hanging curtains, to wrinkled raisins. In materials science, wrinkling has historically been viewed as a failure mechanism — the sudden transition of a flat surface to a wavy one under some form of critical load[3]. The potential of wrinkling as a method for creating wavy surface patterns, particularly at the microscale, was first demonstrated[4] in 1998 for thin, stiff coatings attached to soft substrates. Wrinkling has since been pursued as a versatile means of generating ordered surface patterns[5] and of harnessing surface topography for a wide range of controllable and tunable applications, such as the measurement of material properties[6], wetting[7], adhesion[7], photonics[8] and electronics[9].

The authors' wrinkled microparticles are produced by drying precursors consisting of a soft polymeric core coated with a stiff shell of silica — a process similar to the way that raisins are formed. The core (which corresponds to the soft pulp in a grape) shrinks as the microparticles dry, whereas the shell (corresponding to the grape skin) does not. This generates excess surface area for the shell, which wrinkles spontaneously to accommodate the shrinking core. Previous work[10] studying the wrinkling of thin coatings on soft spherical substrates reported labyrinth patterns similar to those described in the current paper. But Bae and colleagues are among the first to exploit the random and heterogeneous features in disordered wrinkling patterns for applications[8].

The ridged micropatterns in Bae and co-workers' particles are similar to human fingerprints because their major defect features belong to two types of minutia: ridge ending and ridge branching (Fig. 1). When the authors examined the locations and density of minutiae of hundreds of wrinkled microparticles that were generated under identical conditions, they were surprised to find no duplication of patterns — unlike the repeatable wrinkling patterns produced in much of the previously reported work[4–7]. Although the wavelength of the labyrinth ridges is deterministic (mechanistically predictable by theory), the meandering structure of each labyrinth is random and contains minutiae that can function as identifiers.

The researchers characterized the minutiae using the same technique that is used to read conventional fingerprints. They found that the artificial microfingerprints on the particles contain more-randomly oriented minutiae than human fingerprints, and could therefore offer an even higher level of security for identification applications. Furthermore, the wrinkling patterns lock in once they are formed because the silica coating is inelastic, and they survive extreme conditions — such as temperatures of 200 °C and repeated swelling and shrinking in ethanol.

Bae *et al.* report that the number of ridge defects per unit area is inversely proportional to the size of the characteristic wrinkle wavelength, and that more minutiae per surface area can be created using thinner silica coatings, thus enabling the security level of identifier particles to be controlled. Moreover, they show that microfingerprints can be generated on particles that have complex geometries, such as the shapes of letters of the alphabet. This could help to categorize particles to enable their efficient identification by comparison with patterns recorded in databases, and thereby allow a large variety and number of products to be labelled with identifiers. The authors also show that the microscale wrinkling patterns can be decoded using a technique called confocal laser scanning microscopy when particles are attached to products.

The new findings open up the possibility of using defect information in labyrinth wrinkling patterns as unique identifiers for

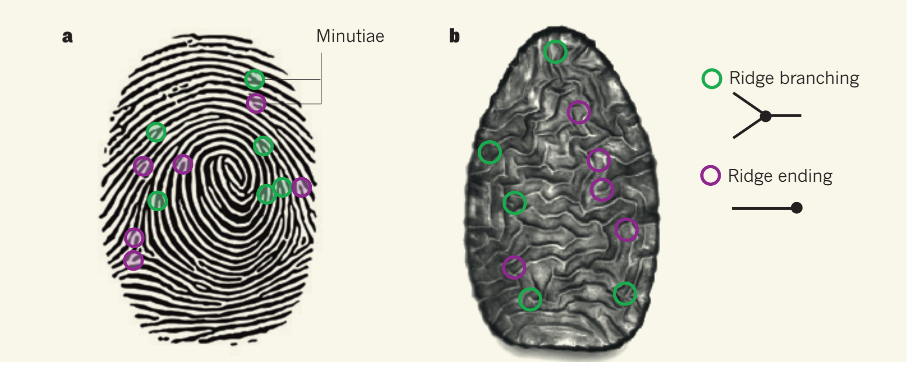


Figure 1 | Similarities between fingerprints and surface wrinkle patterns of microscale particles. a, Human fingerprints are unique labyrinth surface patterns characterized by two types of feature (minutia): ridge ends (purple) and ridge branches (green). **b,** Bae *et al.*[2] report that spontaneously generated labyrinth surface patterns of wrinkles on microparticles are also unique and exhibit similar minutiae. (Images adapted from ref. 2.)

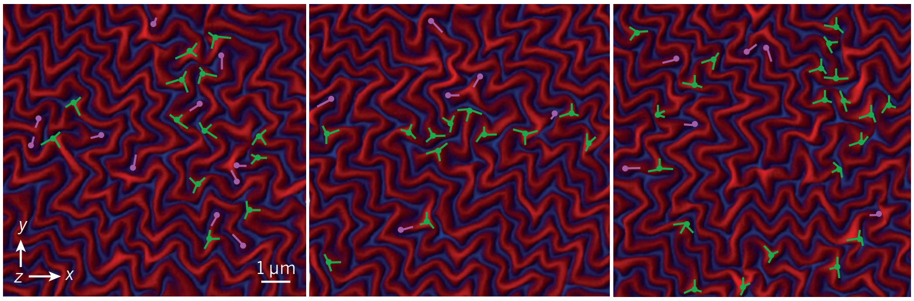


Figure 2 | Sensitivity of wrinkle patterns in films to geometrical imperfections. These computational simulations (conducted by the News & Views authors using the finite element method) depict the wrinkle patterns that form from three systems in which a 250-nanometre-thick film on a planar soft substrate is subjected to the same 'equi-biaxial' compression. Each starting system contained a different thickness imperfection corresponding to just a 0.1% variation of the film's thickness at a few random locations. These seemingly negligible differences result in three distinct wrinkled patterns with different distributions of ridge ends (purple) and ridge branches (green).

encryption, identification and security-check applications. They also raise several questions and suggest future research opportunities. For example, can we better understand and harness the irreproducibility? And what factors govern the randomness?

One possibility is that the observed labyrinth patterns are sensitive to small geometrical or material imperfections in the coating. Similar disordered labyrinth patterns have been observed in thin film coatings on planar soft substrates under 'equi-biaxial' loading[4], and so we conducted some quick numerical simulations of three such systems that differ only by the presence of a few randomly placed imperfections (at which the thickness is 0.1% thicker than at other, perfect, locations). As shown in Figure 2, our simulations reveal that these almost identical systems containing random, negligible geometrical imperfections render three different labyrinth wrinkling patterns under the same loading conditions. It remains to be seen whether the same is true for Bae and colleagues' curved systems. The wrinkle patterns on planar substrates are also unique, and could therefore also be applicable for identification applications.

Widespread application of the artificial microfingerprints might currently be limited, because decoding them requires the use of a specialized high-resolution microscope. However, one could use the same wrinkling technique at larger length scales, which would enable patterns to be read more easily (as human fingerprints are). Nevertheless, the availability of irreproducible, random wrinkling patterns at the microscale opens up many opportunities for encryption, and so a simple, portable technique for reading the micro-minutiae information should be developed. ■

Jie Yin *is in the Department of Mechanical Engineering and at the Temple Materials Institute, Temple University, Philadelphia, Pennsylvania 19122, USA.* **Mary C. Boyce** *is in the Fu Foundation School of Engineering and Applied Science, Columbia University, New York, New York 10027, USA.*
e-mails: jieyin@temple.edu; boyce@columbia.edu

1. de Risi, V. *Geometry and Monadology: Lebniz's Analysis Situs and Philosophy of Space* (Birkhäuser, 2007).
2. Bae, H. J. *et al. Adv. Mater.* http://dx.doi.org/10.1002/adma.201405483 (2015).
3. Timoshenko, S. P. & Gere, J. M. *Theory of Elastic Stability* 2nd edn (McGraw-Hill, 1961).
4. Bowden, N., Brittain, S., Evans, A. G., Hutchinson, J. W. & Whitesides, G. M. *Nature* **393,** 146–149 (1998).
5. Genzer, J. & Groenewold, J. *Soft Matter* **2,** 310–323 (2006).
6. Chung, J. Y., Nolte, A. J. & Stafford, C. M. *Adv. Mater.* **23,** 349–368 (2011).
7. Yang, S., Khare, K. & Lin, P.-C. *Adv. Funct. Mater.* **20,** 2550–2564 (2010).
8. Kim, J. B. *et al. Nature Photon.* **6,** 327–332 (2012).
9. Rogers, J. A., Someya, T. & Huang, Y. *Science* **327,** 1603–1607 (2010).
10. Cao, G., Chen, X., Li, C., Ji, A. & Cao, Z. *Phys. Rev. Lett.* **100,** 036102 (2008).

This article was published online on 1 April 2015.

METABOLISM

Growth in the fat lane

Analysis of endothelial cells, which are involved in blood-vessel formation, unexpectedly reveals that proliferation in this cell type depends on fatty-acid oxidation to support DNA synthesis. SEE ARTICLE P.192

ROBERT A. EGNATCHIK & RALPH J. DEBERARDINIS

The formation of blood vessels requires the migration and proliferation of a vascular cell type called endothelial cells. Most cells assimilate the biomass and energy required for proliferation by converting abundant carbon sources such as sugars and amino acids into macromolecular building blocks. But on page 192 of this issue, Schoors *et al.*[1] report that the oxidation of fatty acids to acetyl-CoA molecules generates an unexpected carbon source that is required in endothelial cells to produce nucleotides for DNA synthesis. Blocking fatty-acid oxidation suppresses endothelial-cell proliferation and protects mice from a common form of vision loss called retinopathy of prematurity, which is caused by uncontrolled blood-vessel formation.

Intermediary metabolism — the process by which cells produce and consume energy — is akin to a network of roads that facilitate traffic flow. The overall function of the network relies on a finite set of major intersections at which pathways meet, before diverging to feed sub-networks. Acetyl-CoA is positioned at one of the most complex of these intersections. This metabolite is produced from the degradation of carbohydrates, lipids and proteins, and feeds carbon into the tricarboxylic acid cycle, a major hub for both energy formation and macromolecular synthesis. Furthermore, because acetyl-CoA is the substrate for acetylation, a molecular modification to DNA-associated proteins that is typically associated with active transcription, it is intimately linked with gene expression and the decision to enter a proliferative state[2,3], thereby tethering intermediary metabolism to cell function.

During angiogenesis (the formation of blood vessels), endothelial cells respond to extracellular signals by activating a complex program of proliferation and migration. The energy sources in angiogenic endothelial cells have been characterized[4], but the sources of acetyl-CoA, and their role in angiogenesis, have remained unknown. In some energetically demanding tissues, such as the heart and skeletal muscle, oxidation of long-chain fatty acids provides a rich source of acetyl-CoA. This pathway involves a repeating sequence of reactions in organelles called mitochondria, with each round generating reducing equivalents to produce energy and liberating two carbons from the fatty acid as acetyl-CoA.

To study the role of fatty-acid oxidation (FAO) in the endothelial cells of mice, Schoors *et al.* inactivated carnitine palmitoyltransferase 1A (CPT1A), an enzyme required for mitochondria to import long-chain fatty acids. This rendered endothelial cells unable to proliferate, impairing vessel sprouting and angiogenesis both in cultured cells and in the mouse retina.

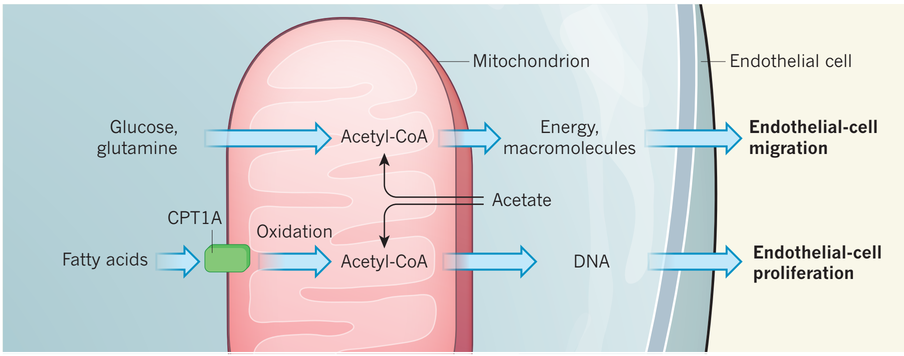


Figure 1 | Metabolic pathways in blood-vessel formation. The formation of blood vessels depends on the proliferation and migration of endothelial cells — processes that require production of the metabolite acetyl-CoA from acetate in organelles called mitochondria. Conversion of glucose, glutamine and other nutrients into acetyl-CoA is required for the production of energy and macromolecules, both of which promote endothelial-cell migration. Schoors *et al.*[1] demonstrate that the production of acetyl-CoA from oxidation of fatty acids, which are imported into mitochondria by the enzyme CPT1A, is required for DNA synthesis and endothelial-cell proliferation. They also report that such fatty-acid oxidation is not required for energy production and cell migration.

Surprisingly, despite the well-known role of FAO in energy production, the authors found that the amount of energy generated by endothelial cells was not altered when CPT1A was absent. Instead, several metabolites produced from acetyl-CoA became depleted, particularly deoxyribonucleotide triphosphates (dNTPs), the building blocks for DNA. Consistent with a specific role for FAO in supporting DNA synthesis, CPT1A loss did not impair the production of either proteins or RNA. Furthermore, providing endothelial cells with dNTPs or acetate, an acetyl-CoA precursor, completely reversed the effects of CPT1A loss on proliferation. Endothelial cells were unusual in this respect; most other proliferating cells, including cancer cell lines, did not use fatty acids as a major carbon source for DNA synthesis.

These findings are of interest for several reasons. First, the requirement for FAO in supporting DNA synthesis would have been difficult to predict on the basis of existing work in proliferative-cell metabolism. Research in cancer-cell metabolism[5] has emphasized the role of glucose and glutamine in feeding the acetyl-CoA pool during growth, and recent work[6] suggests that acetate is also a source of acetyl-CoA in tumours. But even though some tumours require FAO to survive[7], the fact that FAO provides carbon for DNA synthesis was unexpected.

Second, the work strikingly demonstrates divergent metabolic requirements for the two major activities of endothelial cells during angiogenesis: migration and proliferation (Fig. 1). Endothelial cells that lack CPT1A fail to proliferate normally, but their migration is unperturbed. Thus, FAO has a more specialized role in endothelial cells than does glucose metabolism, which these authors previously demonstrated was required for both migration and proliferation[4].

It is to be hoped that Schoors and colleagues' study will stimulate more work to understand this unusual form of metabolic specialization. Given the availability of multiple carbon sources for acetyl-CoA and the tricarboxylic acid cycle, it is intriguing that endothelial cells cannot compensate for CPT1A loss. Does FAO establish a metabolically distinct acetyl-CoA pool that is preferentially channelled towards dNTP synthesis? Additional sophisticated metabolic flux studies, similar to those the authors perform in the paper, might help to answer this question.

An alternative, but not mutually exclusive, explanation is that FAO provides metabolic benefits beyond supplying the cell with carbon. A striking finding from the paper is that CPT1A loss reduces dNTP levels without affecting the precursors of RNA, ribonucleotide triphosphates (rNTPs), perhaps providing a clue to the exquisite requirement of these cells for FAO. Converting rNTPs to dNTPs requires reducing equivalents, which are produced in abundance by FAO. In other models of cell proliferation, conditions that stimulate FAO at the expense of fatty-acid synthesis improve the overall availability of reducing equivalents[8]. It is therefore possible that subtle redox changes contribute to the selective depletion of dNTPs in endothelial cells lacking CPT1A.

The metabolic dependency unveiled by this study has therapeutic potential for diseases that are associated with abnormal endothelial-cell proliferation. The authors demonstrate that systemic FAO blockade with a chemical inhibitor alleviates excessive angiogenesis in a mouse model of retinopathy of prematurity. This disease affects more than 50% of newborn babies with extremely low birth weights worldwide, and is a major source of long-term impairment of visual function. It will be interesting and important to determine whether other forms of abnormal angiogenesis, including the neovascularization of tumours, also require FAO. ■

Robert A. Egnatchik *and* **Ralph J. DeBerardinis** *are at the Children's Medical Center Research Institute, UT Southwestern Medical Center, Dallas, Texas 75390–8502, USA.*
e-mail: ralph.deberardinis@utsouthwestern.edu

1. Schoors, S. *et al. Nature* **520,** 192–197 (2015).
2. Cai, L., Sutter, B. M., Li, B. & Tu, B. P. *Mol. Cell* **42,** 426–437 (2011).
3. Wellen, K. E. *et al. Science* **324,** 1076–1080 (2009).
4. De Bock, K., Georgiadou, M. & Carmeliet, P. *Cell Metab.* **18,** 634–647 (2013).
5. Hensley, C. T., Wasti, A. T. & DeBerardinis, R. J. *J. Clin. Invest.* **123,** 3678–3684 (2013).
6. Comerford, S. A. *et al. Cell* **159,** 1591–1602 (2014).
7. Caro, P. *et al. Cancer Cell* **22,** 547–560 (2012).
8. Jeon, S.-M., Chandel, N. S. & Hay, N. *Nature* **485,** 661–665 (2012).

This article was published online on 1 April 2015.

NUCLEAR CHEMISTRY

Lawrencium bridges a knowledge gap

The heavy element lawrencium is available in only tiny quantities. Measurement of one of its atomic properties was thus an experimental challenge, but indispensably validates theoretical models of heavy elements. **SEE LETTER P.209**

ANDREAS TÜRLER

Accurately measuring the chemical properties of elements that have an atomic number (Z) greater than 100 is exceedingly difficult because of the limited availability of these elements. But by synthesizing single, short-lived atoms of lawrencium ($Z = 103$) in nuclear fusion reactions, Sato *et al.*[1] (page 209 of this issue) have now measured this element's first ionization potential — the energy required to remove a single electron from an uncharged gaseous atom. This allowed them to compare the experimental value with theoretical calculations and to bridge a gap in our knowledge of the chemistry of the actinide elements, of which lawrencium is the final member in the periodic table.

KAZUAKI TSUKADA

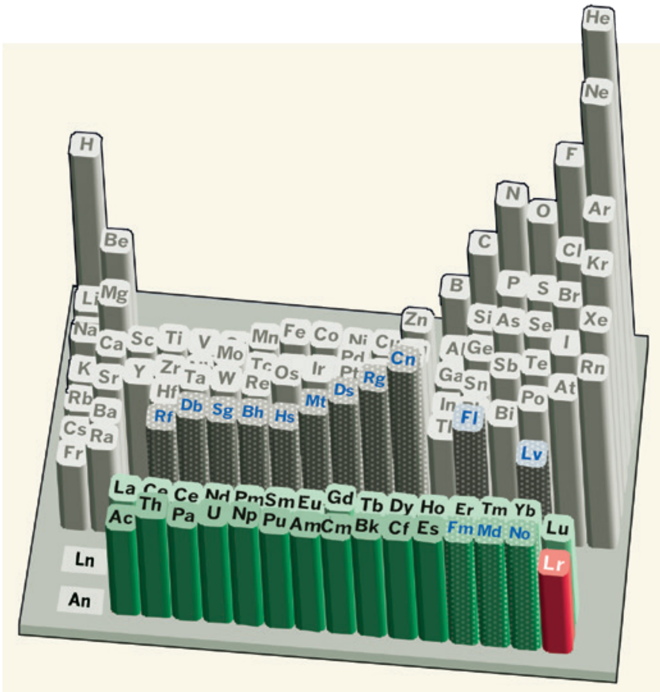


Figure 1 | Periodic table of the elements. Bar sizes represent the energy of the first ionization potential for each element — the energy required to remove a single electron from an uncharged gaseous atom. Sato *et al.*[1] have measured the first ionization potential for lawrencium (Lr; red). The binding energy of the least-bound valence electron of Lr, which can be determined from its ionization potential, is the lowest of those of all the elements, except for group-1 elements heavier than sodium (Na). The first ionization potentials for the elements with symbols depicted in blue have not been experimentally determined. Ln, lanthanides; An, actinides.

Einstein's theory of relativity is perhaps best known for its pivotal role in high-energy physics and astrophysics, but it also has implications for the properties of atoms. The higher the nuclear charge of an atom, the faster the orbiting electrons move, attaining velocities corresponding to a considerable fraction of the speed of light. The relativistic mass of the electrons therefore increases, causing certain inner electron orbitals to contract and stabilize, and thus increasing the shielding of the nuclear charge. This in turn causes other orbitals to expand and destabilize. In addition, 'spin–orbital' interactions become very large for electrons in heavy elements, changing the energy levels of certain orbitals and causing them to split in response to relativistic effects. Theoretical chemists must thus include relativity in their calculations of atomic properties — indeed, this is the only way to explain why gold has its characteristic yellow colour[2] or why mercury is a liquid at room temperature[3].

Experimental study of the superheavy elements that exhibit the greatest relativistic effects is difficult. The heaviest element currently reported has an atomic number of 118, which makes it a member of the noble gases, below xenon and radon in the periodic table. All the other elements in the seventh row of the table (which includes element 118) have also been synthesized. Elements heavier than fermium ($Z=100$) can be produced only as single atoms in nuclear fusion reactions at heavy-ion accelerators, and even the longest-lived isotopes have short half-lives, ranging from several hours to less than one millisecond. The half-life of the lawrencium isotope made by Sato *et al.* is 27 seconds.

The authors synthesized lawrencium by firing a beam of energetic boron-11 ions at a target of californium-249 atoms, producing one lawrencium atom every few seconds in a nuclear fusion reaction. The lawrencium atoms recoil from the californium target because of momentum transfer during fusion, and so the authors captured them by using a helium gas atmosphere seeded with aerosol particles of cadmium iodide — the atoms rapidly diffused to the surface of the aerosol particles.

The helium–aerosol mixture was then passed through a capillary to an ionization site, efficiently transporting the atoms; the ionization site consisted of a metal surface heated to 2,700 or 2,800 kelvin. The researchers measured the efficiency of the ionization of lawrencium atoms on the surface by extracting the ions and sending them through a mass analyser. Cadmium iodide evaporates completely at this temperature, and so did not interfere with the surface ionization process.

From their measurements, Sato and co-workers deduced that the first ionization potential of lawrencium is $4.96^{+0.08}_{-0.07}$ electronvolts. This compares favourably with the calculated value of 4.963(15) eV, which is also reported by the authors in the current paper. Their calculation accounts not only for relativistic effects, but also for minor contributions from quantum electrodynamic effects. These latter effects cause an extra shift in orbital energy levels because of the interaction of electrons with their own electromagnetic radiation field and because of the magnetostatic interaction between electrons.

Lawrencium thus has the lowest first ionization potential of all the lanthanide and actinide elements (Fig. 1). This confirms the proposed picture of the arrangement of the element's outer electrons: the 5*f* orbitals are filled with 14 electrons; a relativistically stabilized 7*s* orbital contains two electrons; and an additional, weakly bound electron resides in the $7p_{1/2}$ valence orbital.

About 10^{12} atoms were needed to determine the first ionization potentials of elements in the past, but, with their spectacular experiment, Sato *et al.* have lowered the amount of material needed by many orders of magnitude.

Does this mean that we will now be able to use this technique to accurately measure all first ionization potentials for superheavy elements up to element 118? Unfortunately not. The ionization potentials for the elements rutherfordium ($Z=104$) to copernicium ($Z=112$) are expected[4] to rise from 6 eV to a staggering 12 eV, the highest value for any known metal. Surface ionization techniques are sadly inefficient for ionization potentials of this magnitude.

Furthermore, Sato and colleagues benefited from being able to generate a few thousand atoms of lawrencium, but for heavier elements it is currently possible to generate only a few atoms — about 1 per week for element 118 (ref. 5). Major improvements in experimental techniques would therefore be needed to study those elements, especially because they are radioactive and have short half-lives. Ideally, an atom (or ion) should be confined in a trap, and several measurements of its properties taken before it decays. Ion traps for nobelium ($Z=102$) and lawrencium have been reported[6], but not for heavier elements. Much ingenuity will be required to push the limits beyond lawrencium, although there is certainly hope that this will become possible.

In the meantime, spectacular developments in relativistic quantum theory and computational algorithms have allowed highly accurate calculations to be made of the basic atomic properties of the heaviest elements. This is just as well because, probably for years to come, theoretical investigations will be the only way to learn more about the chemical properties of some superheavy elements. Sato and co-workers' experimental validation of the predictive power of theoretical methods is therefore highly reassuring. ■

Andreas Türler *is at Paul Scherrer Institute, Villigen PSI, and in the Department of Chemistry and Biochemistry and the Albert Einstein Center for Fundamental Physics, University of Bern, CH-3012 Bern, Switzerland.*
e-mail: andreas.tuerler@psi.ch

1. Sato, T. K. *et al. Nature* **520,** 209–211 (2015).
2. Pyykkö, P. *Inorg. Chim. Acta* **358,** 4113–4130 (2005).
3. Calvo, F., Pahl, E., Wormit, M. & Schwerdtfeger, P. *Angew. Chem. Int. Edn* **52,** 7583–7585 (2013).
4. Pershina, V. in *The Chemistry of Superheavy Elements* 2nd edn (eds Schädel, M. & Shaughnessy, D.) 135–240 (Springer, 2014).
5. Oganessian, Yu. Ts. *et al. Phys. Rev. C* **74,** 044602 (2006).
6. Minaya Ramirez, E. *et al. Science* **337,** 1207–1210 (2012).

NEUROSCIENCE

Binge drinking and brain stress systems

The discovery that interactions between two stress-associated neuropeptide hormones regulate binge drinking in mice provides insight into the biological mechanisms that underpin the bingeing cycle.

RICHARD M. O'CONNOR & PAUL J. KENNY

Many individuals who are normally restrained in their alcohol use will occasionally overindulge, but when alcohol bingeing becomes a regular occurrence, the health and social consequences can be disastrous. Alcohol has mood-enhancing and stress-relieving qualities, but it can also engage stress and anxiety systems in the brain[1], particularly during binge-drinking episodes[2]. These emotional states can be transiently alleviated through further alcohol use, which can lead to a binge-drinking cycle. But precisely how and where in the brain this cycle is regulated is unclear. Writing in *Nature Neuroscience*, Pleil and colleagues[3] report that neuropeptide Y hormone, acting in a brain region called the bed nucleus of the stria terminalis (BNST), suppresses binge drinking in mice by inhibiting the activity of neurons that secrete the stress-related hormone corticotropin-releasing factor.

Neuropeptide Y (NPY) and corticotropin-releasing factor (CRF) are the yin and yang of brain stress systems. Transmission of NPY across the synaptic junctions between adjacent neurons reduces stress and anxiety and promotes feelings of calm[4], whereas transmission of CRF has the opposite effect[5]. These hormones also regulate alcohol consumption, with NPY decreasing[6] and CRF increasing[7] alcohol intake in rodents. Often, receptor proteins for both neuropeptides are located in the same brain regions. The BNST contains an abundance of NPY and CRF receptors, and is a core component of the extended amygdala, a group of closely related brain structures that coordinate responses to stress and reward[8]. Moreover, the BNST regulates stress-induced drug-seeking behaviours[9].

Pleil and colleagues therefore tested the hypothesis that NPY and CRF may interact in the BNST to control binge drinking. First, they investigated the effects of pharmacologically manipulating a type of NPY receptor called Y1R, which inhibits neuronal excitability by modulating activation of the G_i signalling pathway. Activation of Y1R in the BNST of mice decreased binge-like alcohol drinking, and blockade had the opposite effect. Next, Pleil *et al.* dissected the cellular mechanisms by which Y1R exerts these effects. They found that activation of Y1R increased miniature inhibitory postsynaptic currents (mIPSCs), which reflects an increase in release or activity of the inhibitory neurotransmitter molecule GABA. Y1R activation increased the frequency but not the amplitude of mIPSCs — a phenomenon usually interpreted as an increase in the release of GABA from presynaptic neurons. This is a curious observation, given that Y1Rs are located postsynaptically.

How can a postsynaptic receptor act through a presynaptic mechanism? Pleil *et al.* found that the stimulatory effects of Y1R on mIPSC frequency could be abolished either by inhibiting G_i-mediated signalling in postsynaptic neurons, or by disrupting $GABA_A$-receptor insertion into the plasma membrane of postsynaptic neurons, where they must be embedded to be activated. These data point to an unusual phenomenon whereby Y1R activation boosts the ability of GABA to generate inhibitory currents in postsynaptic neurons.

Being nocturnal, mice drink alcohol primarily in the dark phase of the light–dark cycle. Pleil and colleagues could induce Y1R-stimulated increases in mIPSCs only during the light phase of the light–dark cycle in alcohol-naive mice, presumably when natural levels of Y1R activity are low. By contrast, mIPSC increases occurred in both light and dark phases in mice with a history of alcohol bingeing, hinting at alterations in the Y1R system. Accordingly, alcohol-experienced mice demonstrated increased Y1R expression in the BNST compared with those that had never drunk alcohol. Importantly, the authors achieved the same results in experiments on rhesus monkeys. Thus, it seems that the inhibitory influence of Y1R in the BNST is subject to circadian control, but that this control breaks down following episodes of binge drinking.

The responsiveness of individual neurons to the inhibitory effects of NPY varied considerably, indicating that different neuronal types might be variably responsive. Drawing on the idea of NPY and CRF as the yin and yang of stress, the authors hypothesized that NPY in the BNST might directly inhibit CRF-producing neurons. To visualize CRF neurons in the BNST, they engineered mice such that CRF-expressing neurons produced a fluorescent protein. The authors found that only CRF-expressing, fluorescing BNST neurons were inhibited by Y1R stimulation. Moreover, the mIPSCs detected in these neurons were of smaller amplitude than normal, suggesting that Y1R stimulation might alter the composition of $GABA_A$ receptors in CRF neurons (Fig. 1).

Finally, the authors investigated whether Y1R-mediated inhibition of CRF neurons could alter binge alcohol drinking. Using

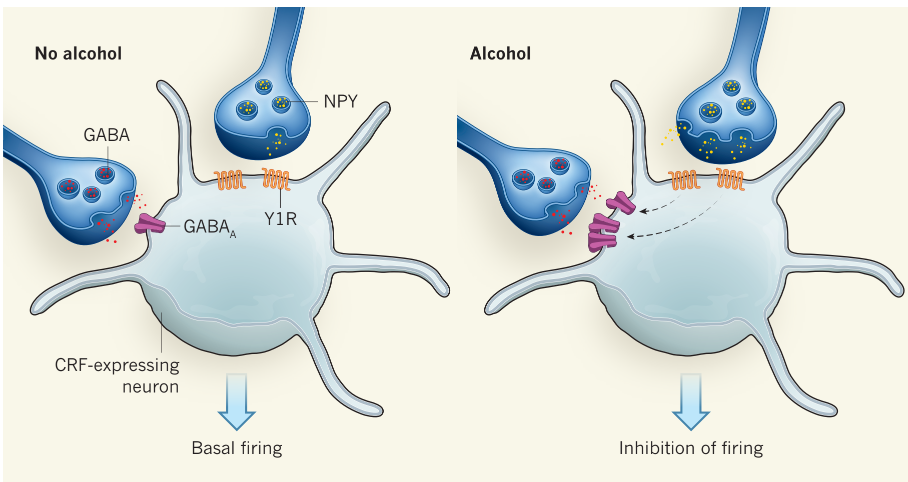


Figure 1 | Control of bingeing patterns. Pleil *et al.*[3] report that, in mice, neurons expressing the hormone neuropepetide Y (NPY) in a brain region called the bed nucleus of the stria terminalis signal to a type of NPY-receptor protein called Y1R in another group of neurons, which express the hormone corticotropin-releasing factor (CRF). In addition, neurons that express the neurotransmitter GABA signal to $GABA_A$ receptors on the membranes of CRF-expressing neurons. Activation of Y1R and $GABA_A$ inhibits CRF-neuron firing. In the absence of alcohol, low levels of Y1R activity mean that CRF neurons fire at basal levels. However, when mice drink alcohol, NPY signalling through Y1R increases. The authors demonstrate that this increases the number of $GABA_A$ receptors in the CRF-neuron membrane, resulting in decreased neuronal firing and so opposing the effects of alcohol consumption.

sophisticated 'chemogenetic' tools, they demonstrated that inhibition of CRF neurons in the BNST suppressed binge drinking and anxiety-like behaviours. Conversely, excitation of CRF neurons blocked the inhibitory effects of Y1R activation on binge drinking.

Pleil and colleagues' work identifies the BNST as a key neuroanatomical hub in which NPY and CRF interact to regulate binge drinking. The findings show that alcohol can induce changes in NPY signalling that might protect against the development of binge patterns of drinking, suggesting that shifts in the balance of power between NPY and CRF might change an individual's vulnerability to bingeing cycles. Moreover, medications that shift this balance in favour of NPY could be a good approach to treating binge drinking.

Are NPY and GABA co-released from the same neurons, and precisely how does Y1R activation increase $GABA_A$ function in CRF neurons? An understanding of these processes might help to expose the molecular mechanisms that regulate synaptic plasticity in the BNST and beyond. The authors' findings also raise the intriguing issue of whether NPY–CRF interactions in the BNST are homeostatic, such that CRF also influences NPY signalling. If so, it might be that changes in CRF activity can in turn suppress NPY activity to precipitate the binge-drinking cycle. Finally, CRF and NPY are thought to have key roles in regulating mood. It is therefore tempting to speculate that NPY–CRF interactions can influence the depression and anxiety that often accompany alcohol-use disorders. ■

Richard M. O'Connor *and* **Paul J. Kenny** *are in the Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, New York, New York 10029–6574, USA.*
e-mail: paul.kenny@mssm.edu

1. Heilig, M. & Koob, G. F. *Trends Neurosci.* **30,** 399–406 (2007).
2. Pohorecky, L. A. *Neurosci. Biobehav. Rev.* **5,** 209–229 (1981).
3. Pleil, K. E. *et al. Nature Neurosci.* http://dx.doi.org/10.1038/nn.3972 (2015).
4. Heilig, M. *Neuropeptides* **38,** 213–224 (2004).
5. Reul, J. M. H. M. & Holsboer, F. *Curr. Opin. Pharmacol.* **2,** 23–33 (2002).
6. Gilpin, N. W. *et al. Biol. Psychiat.* **69,** 1091–1099 (2011).
7. Roberto, M. *et al. Biol. Psychiat.* **67,** 831–839 (2010).
8. Koob, G. F. & Le Moal, M. *Nature Neurosci.* **8,** 1442–1444 (2005).
9. Erb, S. & Stewart, J. *J. Neurosci.* **19,** RC35 (1999).

SOLAR SYSTEM

An incredible likeness of being

Earth and the Moon share many puzzling chemical similarities. New analyses show that the last planet-sized body to hit Earth could have been similar enough to Earth to yield a Moon with an Earth-like composition. SEE LETTER P.212

ROBIN M. CANUP

Some 4.5 billion years ago, our planet grew through a series of violent collisions with other planet-sized bodies, the last of which is thought to have produced the Moon. The impact of a roughly Mars-sized body with Earth can account for the Moon's mass and unusually small iron core, as well as Earth's rapid early spin rate, but it creates a Moon derived mainly from the impacting planet rather than from Earth[1]. It has been thought, mainly on the basis of the observed differences between Earth and Mars, that the impactor's composition would differ substantially from that of Earth[2], and thus would most naturally produce a Moon distinct from Earth in its chemical composition. Instead, Earth and the Moon are nearly chemically identical in many respects. Now, in a paper on page 212, Mastrobuono-Battisti and colleagues[3] estimate a substantial probability — of the order of 20% — that the giant impactor had an Earth-like composition, offering at least a partial solution to the 'isotopic crisis'[4] facing the impact theory.

A key constraint on any hypothesis is that the proportions of different isotopes of a given element vary slightly across samples from the Moon, Earth, Mars and the main asteroid belt that lies between the orbits of Mars and Jupiter. Perhaps the best-studied example involves oxygen (Fig. 1). Earth and the Moon have almost identical oxygen isotope compositions, whereas the difference in oxygen composition between terrestrial rocks and meteorites from Mars or the large asteroid Vesta is considerable[5,6]. Earth and the Moon also share similar isotopic compositions for chromium, silicon, titanium and tungsten, elements whose isotopic abundances vary across meteorites from Mars and the asteroid belt[7].

How could the collision of two large and independently formed planets yield a chemically similar Earth and Moon, when Mars and

NATURE

50 Years Ago

The award of the Order of Merit (the personal gift of the Queen) to Prof. Dorothy Hodgkin makes history in that she is the first woman scientist ever to receive it, and only the second woman to be so honoured by the monarch since the Order was instituted. That she should now be joining Florence Nightingale in this eminence has its parallels and its contrasts. Dorothy Hodgkin is a person of very deep humanitarian interests and concerns … whose work, incidentally, has been carried out against the background of a demanding but happy family life. Her special genius has lain in the unravelling of the structure, by X-ray methods, of complex molecules, and its special timeliness has lain in the fact that she has carried it deep into the realm of biologically important substances (see *Nature* of December 5, 1964). Dorothy Hodgkin received the Nobel Prize in 1964.
From *Nature* 10 April 1965

100 Years Ago

Until the last few years the word margarine was usually associated, in the mind of the British public, with poverty; but now, under the new name of "Nuts and milk," with which advertising enterprise has made us familiar, it is becoming freely used in the kitchen, and is even found on the breakfast table in many households … It is desirable at the outset to emphasise that the margarine industry is essentially scientific in character, and that considerable technical skill is demanded in its manufacture … moreover … edible fats available for the masses at half the price of butter … must be proclaimed as yet another of the achievements of science in the service of man.
From *Nature* 8 April 1915

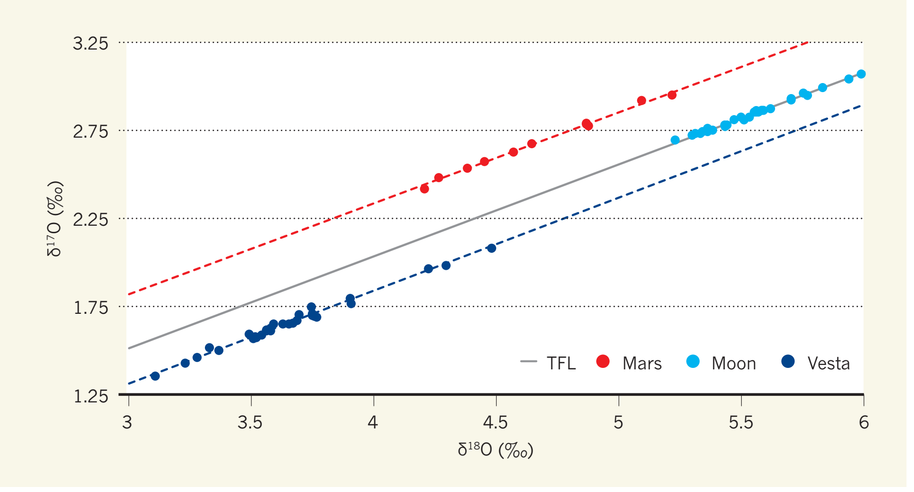


Figure 1 | Oxygen isotopic similarities and differences. The oxygen isotopic compositions of rock samples from Earth and the Moon[12], and from meteorites from Mars[5] and the asteroid Vesta[17], define parallel lines on this graph of the isotopic-abundance ratio of ^{17}O to ^{16}O ($\delta^{17}O$) against that of ^{18}O to ^{16}O ($\delta^{18}O$); the ratios are shown as deviations in parts per thousand (‰) from a standard value. The various lines probably reflect different starting oxygen reservoirs that existed in the pre-planetary disk around the Sun, with the grey line indicating the 'terrestrial fractionation line' (TFL) along which Earth samples fall. The offset in the line for Mars compared with that for Earth is about 0.32‰ (ref. 5), whereas the Earth and Moon lines have an offset of only about 0.01‰ (ref. 6), essentially indistinguishable on the scale of this figure. Mastrobuono-Battisti *et al.*[3] present analyses that explain the Earth–Moon isotopic similarity in the context of a giant impact of a planet-sized body with early Earth. (Figure based on data from refs 5, 12 and 17.)

most meteorites are so different? A canonical Moon-forming impactor contains about 15% of Earth's mass and produces a Moon with more than 80% of its material derived from the impactor[1]. For this to be consistent with the great similarity of Earth and the Moon in oxygen composition, the impactor would need to be nearly as similar to Earth in oxygen isotopic composition as is the Moon. Previous work[2] suggests that this would occur for only a percentage or so of all giant impacts, making the conventional Moon-forming impact an uncomfortably rare event and motivating a search for alternative impact scenarios[2,8,9].

Mastrobuono-Battisti and colleagues used data from simulations[10] that follow the growth of the Solar System's inner planets from an initial distribution of several thousand smaller 'protoplanets' orbiting the Sun. The simulations track merging collisions between the protoplanets that produce a final system of, typically, three or four planets, with the largest comparable in mass to Earth. From these data, the authors extracted the orbital location of each initial protoplanet that contributed to a final planet and its last giant impactor. Following previous work[2], they assumed that the protoplanets had oxygen compositions that varied linearly with distance from the Sun, with a distribution set so that the final third and fourth planets produced by the simulation are as different as Earth and Mars. They then calculated the compositional difference between each planet and its last giant impactor.

They found that about 20% of late giant impactors are close compositional matches to their target planets, which is consistent with an impactor-derived Moon having essentially the same composition as Earth. This is roughly ten times higher than previous estimates. In addition, the predicted compositional difference between a planet and its last giant impactor is found to be substantially smaller than the characteristic differences between the final planets. In other words, the same initial compositional mix of protoplanets implied by the Earth–Mars difference can still be consistent with the Earth–Moon similarity.

Two factors seem to have increased the likelihood of an impactor–planet compositional match. First, Mastrobuono-Battisti *et al.* used more-recent planet-formation simulations[10], each with a tenfold higher resolution (that is, smaller initial protoplanets) than the simulation[11] analysed previously. Trends in the new analysis suggest that, as the growth of the impactor itself is more finely resolved, predicted impactor–planet compositional differences decrease. All other factors being equal, assembling an impactor from a larger number of smaller pieces would tend to leave it on a less-elliptical orbit, perhaps reducing its incorporation of more distant, compositionally distinct material. This suggests that higher resolutions than those of the simulations analysed by the authors might yield still smaller differences, a hypothesis testable by future work. Second, the present work relies on a recently reported difference in oxygen composition between Earth and the Moon, which although still very small[6] (12 ± 3 parts per million), increases the probability of a successful impactor–planet match compared with the previous upper limit[12] (less than 5 parts per million).

Mastrobuono-Battisti and co-workers' findings offer renewed support for the canonical impact by providing a plausible explanation for the compositional similarities between Earth and the Moon, apart from two possible exceptions: silicon and tungsten. Earth and the Moon are enhanced in the heavy isotopes of silicon, which has been attributed to the preferential incorporation of the light isotope of silicon into Earth's core[13]. Because this effect requires high pressures inside an Earth-sized planet, it would not occur in a smaller Mars-sized impactor, implying that a Moon derived from such a body would have a lighter silicon composition, inconsistent with that observed[13]. However, heavy silicon compositions among meteorites that are unlikely to have originated in a massive planet have now been identified[14], suggesting that a Mars-sized impactor heavy in silicon might be possible as well.

A remaining puzzle is tungsten, whose isotopic composition is affected by the nature and timing of the formation of an object's iron core. Studies[15,16] just published find a small difference in tungsten composition between Earth and the Moon, which is probably due to differing proportions of material accumulated by each object after the Moon-forming impact. Although the closeness of the Earth–Moon tungsten compositions could result from a canonical impact[7], it seems a coincidence whose likelihood needs to be quantified. Even with such constraints, the conventional impact may ultimately prove more probable than alternatives that either require rarer types of impact or additional processes, or both[2,8,9]. ■

Robin M. Canup *is at the Southwest Research Institute, Boulder, Colorado 80302, USA.*
e-mail: robin@boulder.swri.edu

1. Canup, R. M. *Icarus* **168,** 433–456 (2004).
2. Pahlevan, K. & Stevenson, D. J. *Earth Planet. Sci. Lett.* **262,** 438–449 (2007).
3. Mastrobuono-Battisti, A., Perets, H. B. & Raymond, S. N. *Nature* **520,** 212–215 (2015).
4. Melosh, H. J. *Phil. Trans. R. Soc. A* **372,** 20130175 (2014).
5. Franchi, I. A., Wright, I. P., Sexton, A. S. & Pillinger, C. T. *Meteorit. Planet. Sci.* **34,** 657–661 (1999).
6. Herwartz, D., Pack, A., Friedrichs, B. & Bischoff, A. *Science* **344,** 1146–1150 (2014).
7. Dauphas, N., Burkhardt, C., Warren, P. H. & Teng, F.-Z. *Phil. Trans. R. Soc. A* **372,** 20130244 (2014).
8. Ćuk, M. & Stewart, S. T. *Science* **338,** 1047–1052 (2012).
9. Canup, R. M. *Science* **338,** 1052–1055 (2012).
10. Raymond, S. N., O'Brien, D. P., Morbidelli, A., Kaib, N. A. *Icarus* **203,** 644–662 (2009).
11. Chambers, J. *Icarus* **152,** 205–224 (2001).
12. Wiechert, U. H. *et al. Science* **294,** 345–348 (2001).
13. Georg, R. B., Halliday, A. N., Schauble, E. A. & Reynolds, B. C. *Nature* **447,** 1102–1106 (2007).
14. Dauphas, N., Poitrasson, F. & Burkhardt, C. *46th Lunar Planet. Sci. Conf.* Abstr. 1417 (2015).
15. Touboul, M., Putchtel, I. S. & Walker, R. J. *Nature* http://dx.doi.org/10.1038/nature14355 (2015).
16. Kruijer, T. S., Kleine, T., Fischer-Gödde, M. & Sprung, P. *Nature* http://dx.doi.org/10.1038/nature14360 (2015).
17. Wiechert, U. H., Halliday, A. N., Palme, H. & Rumble, D. *Earth Planet. Sci. Lett.* **221,** 373–382 (2004).

REVIEW

doi:10.1038/nature14338

Climate change and the permafrost carbon feedback

E. A. G. Schuur[1,2], A. D. McGuire[3], C. Schädel[1,2], G. Grosse[4], J. W. Harden[5], D. J. Hayes[6], G. Hugelius[7], C. D. Koven[8], P. Kuhry[7], D. M. Lawrence[9], S. M. Natali[10], D. Olefeldt[11,12], V. E. Romanovsky[13,14], K. Schaefer[15], M. R. Turetsky[11], C. C. Treat[16] & J. E. Vonk[17]

Large quantities of organic carbon are stored in frozen soils (permafrost) within Arctic and sub-Arctic regions. A warming climate can induce environmental changes that accelerate the microbial breakdown of organic carbon and the release of the greenhouse gases carbon dioxide and methane. This feedback can accelerate climate change, but the magnitude and timing of greenhouse gas emission from these regions and their impact on climate change remain uncertain. Here we find that current evidence suggests a gradual and prolonged release of greenhouse gas emissions in a warming climate and present a research strategy with which to target poorly understood aspects of permafrost carbon dynamics.

In high-latitude regions of Earth, temperatures have risen 0.6 °C per decade over the last 30 years, twice as fast as the global average[1]. This is causing normally frozen ground to thaw[2–4], exposing substantial quantities of organic carbon to decomposition by soil microbes. This permafrost carbon is the remnant of plants and animals accumulated in perennially frozen soil over thousands of years, and the permafrost region contains twice as much carbon as there is currently in the atmosphere[5,6]. Conversion of just a fraction of this frozen carbon pool into the greenhouse gases carbon dioxide (CO_2) and methane (CH_4) and their release into the atmosphere could increase the rate of future climate change[7]. Climate warming as a result of human activities causes northern regions to emit additional greenhouse gases to the atmosphere, representing a feedback that will probably make climate change happen faster than is currently projected by Earth System models. The critical question centres on how fast this process will occur, and recent publications differ in their outlook on this issue. Abrupt releases of CH_4 forecast to cause trillions of dollars of economic damage to global society[8] contrast with predictions of slower, sustained greenhouse gas release that, although substantial, would give society more time to adapt[1,9]. This range of viewpoints is due in part to the wide uncertainty surrounding processes that are only now being quantified in these remote regions.

Here we provide an overview of new insights from a multi-year synthesis of data with the aim of constraining our current understanding of the permafrost carbon feedback to climate, and providing a framework for developing research initiatives in the permafrost region[10,11]. We begin by reviewing new research, much of it published since the Intergovernmental Panel on Climate Change (IPCC)'s Fifth Assessment Report (AR5)[1], on the size of the carbon pool stored in the permafrost region. Synthesis research has enlarged the number of observations in the permafrost region soil carbon pool database tenfold[12], and confirms that tremendous quantities of carbon accumulated deep in permafrost soils are widespread[5,6]. We then discuss new long-term laboratory incubations of these permafrost soils that reveal that a substantial fraction of this material can be mineralized by microbes and converted to CO_2 and CH_4 on timescales of years to decades, which would contribute to near-term climate warming. Initial estimates of greenhouse gas release point towards the potential for substantial emissions of carbon from permafrost in a warmer world, but these could still be underestimates. Field observations reveal that abrupt thaw processes are common in northern landscapes, but our review shows that mechanisms that speed thawing of frozen ground and release of permafrost carbon are entirely absent from the large-scale models used to predict the rate of climate change.

Bringing together this wealth of new observations, we propose that greenhouse gas emissions from warming permafrost are likely to occur at a magnitude similar to other historically important biospheric carbon sources (such as land-use change) but that will be only a fraction of current fossil-fuel emissions. At the proposed rates, the observed and projected emissions of CH_4 and CO_2 from thawing permafrost are unlikely to cause abrupt climate change over a period of a few years to a decade. Instead, permafrost carbon emissions are likely to be felt over decades to centuries as northern regions warm, making climate change happen faster than we would expect on the basis of projected emissions from human activities alone. This improved knowledge of the magnitude and timing of permafrost carbon emissions based on the synthesis of existing data needs to be integrated into policy decisions about the management of carbon in a warming world, but at the same time may help temper the worst fears about the impact of carbon emissions from warming northern high-latitude regions.

Permafrost carbon pool

The first studies that brought widespread attention to permafrost carbon estimated that almost 1,700 billion tons of organic carbon were stored in terrestrial soils in the northern permafrost zone[6,7,13]. The recognition of this vast pool stored in Arctic and sub-Arctic regions was in part due to substantial carbon stored at depth (>1 m) in permafrost, below the traditional zone of soil carbon accounting[14]. Deeper carbon measurements were initially rare, and it was not even possible to quantify the uncertainty for the permafrost carbon pool size estimate. However, important new syntheses continue to report large quantities of deep carbon preserved in permafrost at many previously unsampled locations, and that a substantial fraction of this deep

[1]Center for Ecosystem Science and Society and Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona 86011, USA. [2]Department of Biology, University of Florida, Gainesville, Florida 32611, USA. [3]US Geological Survey, Alaska Cooperative Fish and Wildlife Research Unit, University of Alaska Fairbanks, Alaska 99775, USA. [4]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, 14473 Potsdam, Germany. [5]US Geological Survey, Menlo Park, California 94025, USA. [6]Climate Change Science Institute and Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA. [7]Department of Physical Geography, Stockholm University, 10691 Stockholm, Sweden. [8]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [9]National Center for Atmospheric Research, Boulder, Colorado 80305, USA. [10]Woods Hole Research Center, Falmouth, Massachusetts 02540, USA. [11]Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada. [12]Department of Renewable Resources, University of Alberta, Edmonton, Alberta T6G 2H1, Canada. [13]Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska 99775, USA. [14]Tyumen State Oil and Gas University, Tyumen, Tyumen Oblast 625000, Russia. [15]National Snow and Ice Data Center, Boulder, Colorado 80309, USA. [16]Earth Systems Research Center, Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, New Hampshire 03824, USA. [17]Department of Earth Sciences, Utrecht University, 3584 CD Utrecht, The Netherlands.

permafrost carbon is susceptible to future thaw[15]. The permafrost carbon pool is now thought to comprise organic carbon in the top 3 m of surface soil, carbon in deposits deeper than 3 m (including those within the yedoma region, an area of deep sediment deposits that cover unglaciated parts of Siberia and Alaska[16–18]), as well as carbon within permafrost that formed on land during glacial periods but that is now found on shallow submarine shelves in the Arctic. Recent research has expanded our knowledge considerably while at the same time highlighting remaining gaps in our understanding of this vulnerable carbon pool[19].

Surface carbon

The new northern permafrost zone carbon inventory reports the surface permafrost carbon pool (0–3 m) to be 1,035 ± 150 Pg carbon (mean ± 95% confidence interval, CI)[12,20] (where 1 Pg = 1 billion tons) (Fig. 1a). This estimate supported the original studies while improving precision by increasing the number of deeper (>1 m) sampling locations tenfold. This surface permafrost carbon pool is substantial. The rest of Earth's biomes, excluding the Arctic and boreal regions, are thought to contain 2,050 Pg carbon in the surface 3 m of soil[21]. Even though these northern regions account for only 15% of global soil area, the 0–3 m global soil carbon pool is increased by 50% when fully accounting for the carbon stored deeper in permafrost zone soil profiles.

Deep carbon in yedoma

Processes that accumulate carbon deep into permafrost soils do not stop at 3 m depth, and our previously limited understanding of those deep carbon deposits (>3 m depth) has been improved. In particular, several new estimates have emerged for carbon that accumulated during, and since, the last Ice Age in the yedoma region in Siberia and Alaska[16–18]. These new data support previous findings of relatively high carbon concentrations in permafrost soil at depth, but revised the understanding of total carbon stock by improving the estimates of spatial extent, type of deposit, sediment depth, and ground ice content. These deep, perennially frozen sediments are particularly ice-rich, where ice occupies 50%–80% of the ground volume[22,23]. Although this excess ice does not alter soil carbon concentration, it affects the total carbon inventory contained in a particular volume of soil, decreasing carbon stocks per unit soil volume by 22%–50% compared to previous estimates[24]. Because of the continued difficulty of measuring total ground ice content and total sediment depth, carbon pool estimates for the yedoma region still range by twofold even as new data from this region have accumulated. This region is now thought to contain between 210 ± 70 Pg carbon (ref. 16) and 456 ± 45 Pg carbon (ref. 18), still supporting the original accounts of several hundred billion tons of carbon stored deep in the permafrost even when recalculated with new observations.

Deep carbon outside the yedoma region

While new measurements of deep carbon have been largely focused on the 1.2 million square kilometres of the yedoma region in recent years, other areas in the northern permafrost zone with thick loose sedimentary material may also contain substantial organic carbon pools in permafrost (Fig. 1b). The major Arctic river deltas are now thought to contain 91 ± 39 Pg carbon (95% CI)[12], while carbon contained in the approximately 5 million square kilometres of thick (>5–10 m) sediments overlying bedrock outside the yedoma and river delta regions remain largely unknown. Taking the spatial extent of these poorly known permafrost areas, along with an estimated thickness in the tens of metres (similar to that of yedoma), and average carbon content of a few deep borehole soil samples, there could be an additional deep permafrost carbon pool of 350–465 Pg C outside the yedoma region (calculated using a depth interval of 3–10 m and carbon content of 11–14 kg C m^{-3}, which accounts for ground ice[25]).

Subsea permafrost carbon

Much of the inventory until this point has focused on terrestrial ecosystems where permafrost is currently sustained by cold winter air temperatures. But permafrost also exists below Arctic Ocean continental shelves, in particular the East Siberian Arctic Shelf, the largest and shallowest shelf on Earth. This permafrost is an extension of the terrestrial permafrost that existed during the last Ice Age, but became submerged when sea level rose during the late Pleistocene–Holocene transition, and at the beginning of the Holocene epoch. The shallow shelf area exposed as dry land in the area around Alaska and Siberia during the last Ice Age (<125 m current ocean depth), at almost 3 million square kilometres, is about 2.5 times the size of the current terrestrial yedoma region[16,26]. But the quantity of organic permafrost carbon stored beneath the sea floor is even more poorly quantified than on land and could be lower than it once was[27,28]. Subsea permafrost as a whole has been slowly degrading over thousands of years as relatively warm ocean water has warmed the newly submerged sea floor. Frozen sediments are thickest near the shore, where submergence with seawater occurred more recently than on the outer shelf, which is now underlain by discontinuous, patchy permafrost[29,30]. During this time of thaw, organic carbon was mineralized by microbes within the sediment in low-oxygen conditions that promote the formation of CH_4, reducing the pool of permafrost carbon remaining under the sea.

Taken together, the known pool of terrestrial permafrost carbon in the northern permafrost zone is 1,330–1,580 Pg carbon, accounting for surface carbon as well as deep carbon in the yedoma region and river deltas, with the potential for ~400 Pg carbon in other deep terrestrial permafrost sediments that, along with an additional quantity of subsea permafrost carbon, still remains largely unquantified.

Carbon decomposability

Permafrost carbon stocks provide the basis for greenhouse gas release to the atmosphere, but the rate at which this can happen is also controlled by the overall decomposability of organic carbon. Conceptual models and initial data on decomposability suggested that a portion of permafrost carbon is susceptible to rapid breakdown upon thaw[13,31]. But it has not been clear to what degree this could be sustained on the decade-to-century timescale of climate change, or what degree of variation exists within soils across the vast landscape of the permafrost zone. New research has confirmed that initial rates of permafrost carbon loss are potentially high, but continued observation reported declines in carbon loss rates over time, which might be expected as more labile carbon pools are exhausted[32]. This has highlighted the need for long-term observation under controlled conditions to estimate the potential decomposability of permafrost carbon. New data from a 12-year incubation of permafrost soil from Greenland showed that 50%–75% of the initial carbon was lost by microbial decomposition under aerobic and continuously unfrozen laboratory conditions over that time frame[33]. This experiment, of unprecedented length for permafrost soils compared to typical incubations that might be only weeks to months long[34,35], was then extended geographically in a new synthesis of long-term (>1 year) permafrost zone soil incubations. Soils from across the permafrost region showed similarly high potential for microbial degradation of organic carbon upon thaw in the laboratory, with a wider range of decade-long losses projected to be 1%–76% (Fig. 2a) under laboratory conditions[36].

A major cause of landscape-scale variation in decomposability across soils was linked to the carbon to nitrogen ratio of the organic matter, with higher values leading to more greenhouse gas release. This simple metric (the carbon to nitrogen ratio) is in part illustrated by grouping soils as organic (>20% C) with mean decade-long losses of 17%–34% (lower-to-upper 97.5% CI) and mineral (<20% C) with mean decade-long losses of 6%–13% (Fig. 2a). The metric takes into account the ability of microbes to process permafrost carbon for metabolism by breaking down organic carbon for energy, and to grow by acquiring nutrients such as nitrogen released during the decomposition process. Because carbon and nitrogen are often measured in soil surveys, maps of permafrost carbon pools can then be combined with the findings from laboratory incubations to project potential carbon emission estimates across the permafrost region to determine which regions could be emission hotspots in a warming climate. The location of such potential emission hotspots is expected to be affected by both the total pool of permafrost carbon and the potential for that carbon to be broken

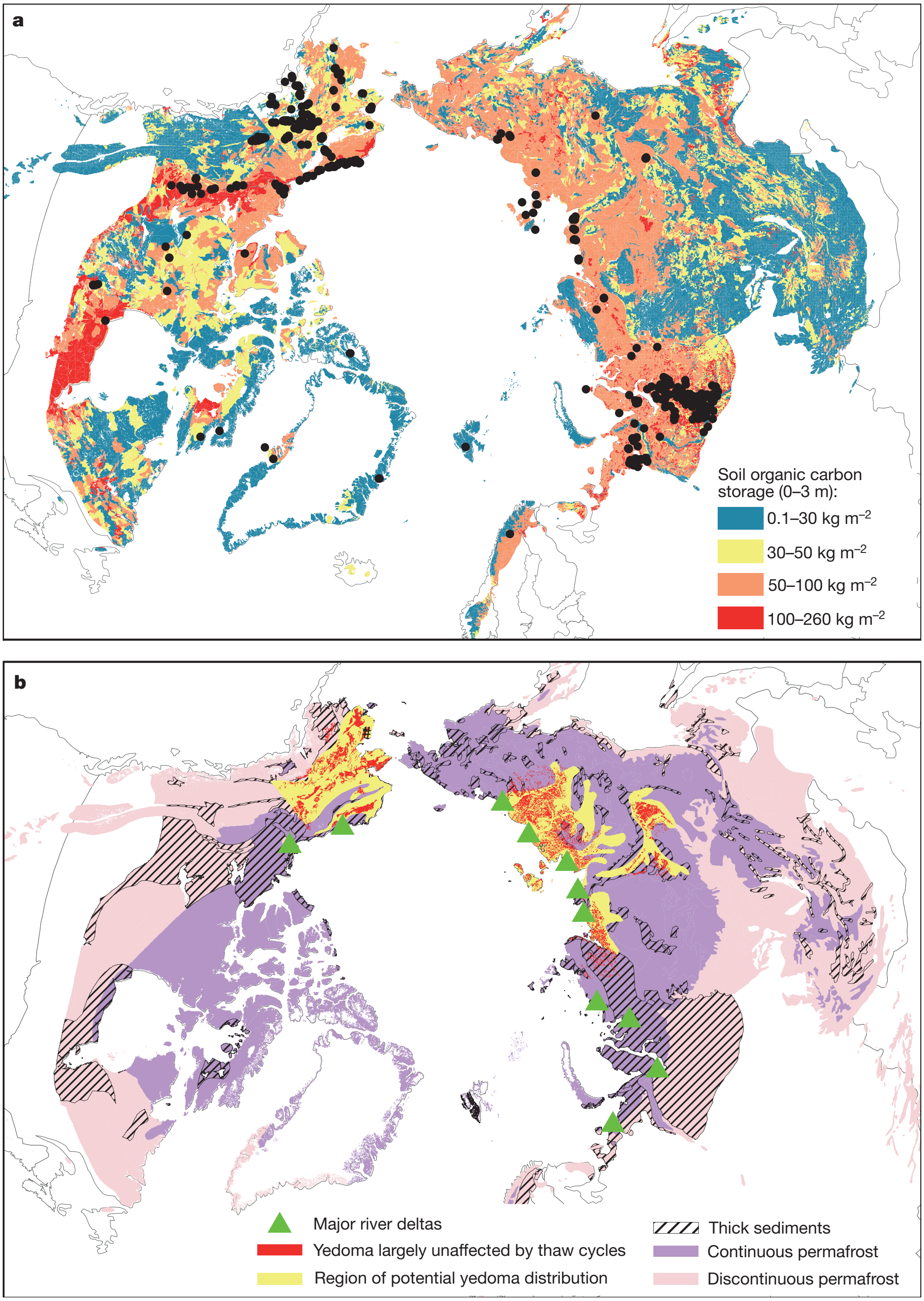


Figure 1 | Soil organic carbon maps. a, Soil organic carbon pool (kg C m^{-2}) contained in the 0–3 m depth interval of the northern circumpolar permafrost zone[12]. Points show field site locations for 0–3 m depth carbon inventory measurements; field sites with 1 m carbon inventory measurements number in the thousands and are too numerous to show. **b**, Deep permafrost carbon pools (>3 m), including the location of major permafrost-affected river deltas (green triangles), the extent of the yedoma region previously used to estimate the carbon content of these deposits[13] (yellow), the current extent of yedoma region soils largely unaffected by thaw-lake cycles that alter the original carbon content[17] (red), and the extent of thick sediments overlying bedrock (black hashed). Yedoma regions are generally also thick sediments. The base map layer shows permafrost distribution with continuous regions to the north having permafrost everywhere (>90%), and discontinuous regions further south having permafrost in some, but not all, locations (<90%)[96].

down by microbes after thaw as controlled by the energy and nutrients contained within the organic matter.

The inherent range of permafrost carbon decomposability across soil types also intersects with environmental conditions, and aerobic decomposition is only part of the story for northern ecosystems. While temperature control over decomposition is implicit when considering permafrost thaw, this region is characterized by widespread lakes, wetlands, and soils waterlogged as a result of surface drainage restricted by underlying permafrost. The lack of oxygen in saturated anaerobic soils and sediments presents another key control over emissions from newly thawed permafrost carbon. Comparing the results from the aerobic permafrost soil incubation synthesis[36] with those from another circumpolar synthesis of anaerobic soil incubations[37]

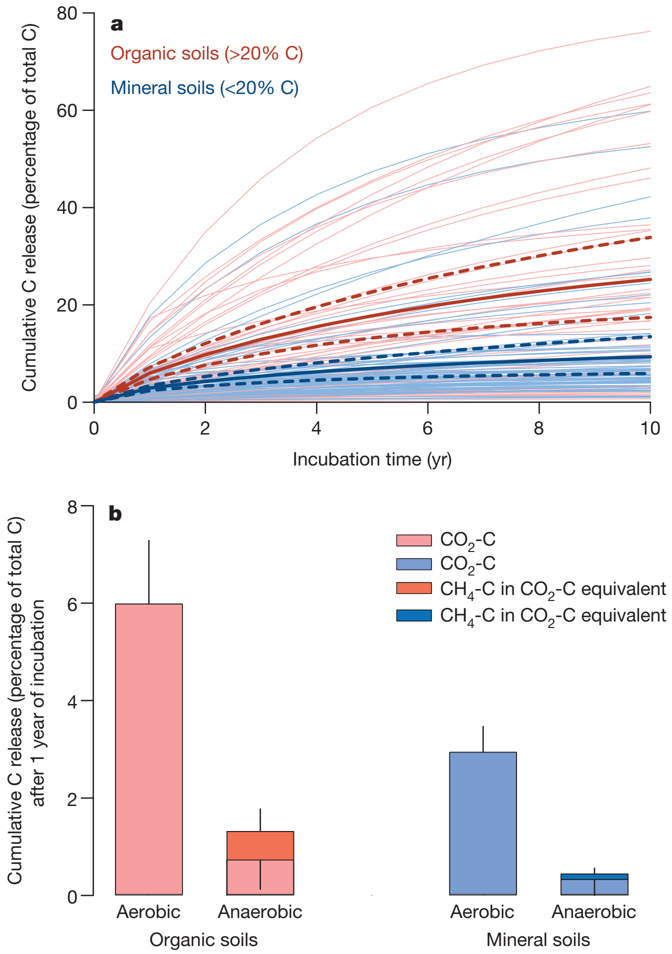


Figure 2 | Potential cumulative carbon release. Data are given as a percentage of initial carbon. **a**, Cumulative carbon release after ten years of aerobic incubation at a constant temperature of 5 °C. Thick solid lines are averages for organic (red, $N = 43$) and mineral soils (blue, $N = 78$) and thin solid lines represent individual soils to show the response of individual soils. Dotted lines are the averages of the 97.5% CI for each soil type. **b**, Cumulative carbon release after one year of aerobic and anaerobic incubations (at 5 °C). Darker colours represent cumulative CH_4-carbon calculated as CO_2-carbon equivalent (for anaerobic soils) on a 100-year timescale according to ref. 38. Positive error bars are upper 97.5% CI for CO_2-carbon and negative error bars are lower 97.5% CI for CH_4-carbon. $N = 28$ for organic soils and $N = 25$ for mineral soils in anaerobic incubations. Aerobic cumulative carbon release is redrawn from ref. 36 and anaerobic cumulative carbon release is calculated based on ref. 37.

shows that cumulative carbon emissions, over an equal one-year incubation time frame, are, on average, 78%–85% lower than those from aerobic soils (Fig. 2b). Specialized microbes release CH_4 along with CO_2 in these environments, and the more potent (that is, it affects climate change more powerfully) greenhouse gas CH_4 in the atmosphere can partially offset a decreased decomposition rate. While mean quantities of CH_4 are 3% (in mineral soils) to 7% (in organic soils) that of CO_2 emitted from anaerobic incubations (by weight of carbon), these mean CH_4 values represent 25% (in mineral soil) to 45% (in organic soil) of the overall potential impact on climate over a 100-year timescale when accounting for CH_4 (ref. 38). Across the mosaic of ecosystems in the permafrost region, controlled laboratory observations brought together here imply that, in spite of the more potent greenhouse gas CH_4, a unit of newly thawed permafrost carbon could have a greater impact on climate over a century if it thaws and decomposes within a drier, aerobic soil as compared to an equivalent amount of carbon within a waterlogged soil or sediment.

Controlled laboratory work is critical for identifying the key mechanisms for potential greenhouse gas release from permafrost carbon, but some important processes are difficult to address with incubation experiments. For example, CH_4 generated from permafrost carbon can be oxidized in aerobic soil layers above the water table and released to the atmosphere as CO_2 instead. This effect can be modified by vegetation, for example, sedge stems acting as pipes provide a pathway for CH_4 to avoid oxidation and to escape to the atmosphere[39]. A synthesis of field CH_4 emission rates showed that sedge-dominated sites had emission rates 2–5 times higher[40], due in part to sedges allowing the physical escape of CH_4, as well as providing more decomposable carbon to the microbial community[41,42]. But even with sedges, it is likely that CH_4 oxidation as a whole would decrease the warming impact of permafrost carbon decomposing in a waterlogged environment compared to what was measured from a laboratory potential. Incubation results, while needing to be interpreted carefully, are useful for scaling the potential of permafrost soils to release greenhouse gases upon thaw, and also for helping to quantify the fraction of soil carbon that is likely to remain relatively inert within the soil after thaw.

Projecting change

A number of ecosystem and Earth system models have incorporated a first approximation of global permafrost carbon dynamics. Recent key improvements include the physical representation of permafrost soil thermodynamics and the role of environmental controls, in particular the soil freeze/thaw state, on decomposition of organic carbon[43–45]. These improved models, which specifically address processes known to be important in permafrost ecosystems but that were missing from earlier model representations, have been key for forecasting the potential release of permafrost carbon with warming, and the impact this would have on the rate of climate change. Model scenarios show potential carbon release from the permafrost zone in the range 37–174 Pg carbon by 2100 under the current climate warming trajectory (Representative Concentration Pathway RCP 8.5), with an average across models of 92 ± 17 Pg carbon (mean ± s.e.) (Fig. 3)[45–52]. Furthermore, thawing permafrost carbon is forecasted to impact global climate for centuries, with models, on average, estimating that 59% of total permafrost carbon emissions will occur after 2100. While carbon releases over these time frames are understandably uncertain, they illustrate the momentum of a warming climate that thaws near-surface permafrost, causing a cascading release of greenhouse gases as microbes slowly decompose newly thawed permafrost carbon. At the scale of these models not all differentiated between CO_2 and CH_4 loss, but expert assessment, a method for surveying expert knowledge, placed CH_4 losses at about 2.3% of total future emissions from the permafrost zone[53,54]. This has the effect, in the expert assessment, of increasing the warming potential of released carbon by 35%–48% when accounting for the more potent greenhouse gas CH_4 over a 100-year timescale.

Within the wide uncertainty of forecasts, some broader patterns are just beginning to emerge. Models vary widely when predicting the current pool of permafrost carbon, which is the source of future carbon emissions in a warmer world. The model average permafrost carbon pool size was estimated at 771 ± 100 Pg carbon (mean ± s.e.), about half as much as the measurement-based estimate, potentially related in part to the fact that models mostly represented carbon to only 3 m depth. A smaller modelled carbon pool could, in principle, constrain forecasted carbon emissions. Normalizing the emissions estimates from the dynamic models by their initial permafrost carbon pool size, 15% ± 3% (mean ± s.e.) of the initial pool was expected to be lost as greenhouse gas emissions by 2100[55]. This decrease in the permafrost carbon pool is similar, but somewhat higher, than the 7%–11% (95% CI) loss predicted by experts[53,54], and the relatively constant fraction across model estimates does hint at the importance of pool size in constraining carbon emissions. However, sensitivity to both modelled Arctic climate change, as well as the responses of soil temperature, moisture and carbon dynamics, are important controls over emissions predictions within these complex models, not pool size alone[44,56,57]. Full diagnosis of the important parameters that regulate the permafrost carbon feedback is not currently possible from the small number of modelling studies that exist, but the estimates do seem to converge on a vulnerable fraction of permafrost carbon that seems to be in line with other approaches.

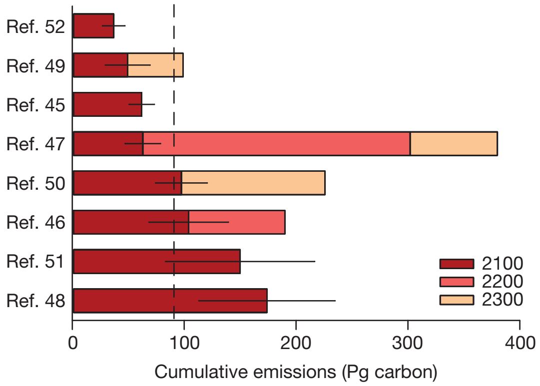


Figure 3 | Model estimates of potential cumulative carbon release from thawing permafrost by 2100, 2200, and 2300. All estimates except those of refs 50 and 46 are based on RCP 8.5 or its equivalent in the AR4 (ref. 97), the A2 scenario. Error bars show uncertainties for each estimate that are based on an ensemble of simulations assuming different warming rates for each scenario and different amounts of initial frozen carbon in permafrost. The vertical dashed line shows the mean of all models under the current warming trajectory by 2100.

These dynamic models also simultaneously assess the countering influence of plant carbon uptake, which may in part offset permafrost carbon release. Warmer temperatures, longer growing seasons, elevated CO_2, and increased nutrients released from decomposing organic carbon may all stimulate plant growth[58]. New carbon can be stored in larger plant biomass or deposited into surface soils[59]. A previous generation of Earth system models that did not include permafrost carbon mechanisms but did simulate changes in plant carbon uptake estimated that the vegetation carbon pool could increase by 17 ± 8 Pg carbon by 2100, with increased plant growth also contributing to new soil carbon accumulation of similar magnitude[60]. The models reviewed here that do include permafrost carbon mechanisms (as well as many of the mechanisms that stimulate plant growth that were used in the previous generation of models) generally indicate that increased plant carbon uptake will more than offset soil carbon emissions from the permafrost region for several decades as climate becomes warmer[45,46,48]. Over longer timescales and with continued warming, however, microbial release of carbon overwhelms the capacity for plant carbon uptake, leading to net carbon emissions from permafrost ecosystems to the atmosphere. Modelled carbon emissions projected under various warming scenarios translate into a range of 0.13–0.27 °C additional global warming by 2100 and up to 0.42 °C by 2300, but currently remain one of the least constrained biospheric feedbacks to climate[1].

Abrupt permafrost thaw

Recent progress towards predicting change in permafrost carbon dynamics focuses mostly on gradual top-down thawing of permafrost. However, increasing evidence from the permafrost zone suggests that abrupt permafrost thaw may be the norm for many parts of the Arctic landscape[17,18,61,62] (Fig. 4). Abrupt permafrost thaw occurs when warming melts ground ice, causing the land surface to collapse into the volume previously occupied by ice. This process, called thermokarst, alters surface hydrology. Water is attracted towards collapse areas, and pooling or flowing water in turn causes more localized thawing and even mass erosion. Owing to these localized feedbacks that can thaw through tens of metres of permafrost across a hillslope within only a few years, permafrost thaw occurs much more rapidly than would be predicted from changes in air temperature alone. This raises the question of whether key complexity is missing from large-scale model projections that are based on first approximations of permafrost dynamics.

Abrupt thaw occurs only at point locations but often causes much deeper permafrost thaw to occur more rapidly. This is in contrast to top-down thawing, which occurs across the entire landscape but affects only the permafrost surface. New regional research is beginning to reveal that a large fraction of permafrost carbon is vulnerable to abrupt thaw. For example, since the end of the last Ice Age, thermokarst thaw-lake cycles have affected 70% of the yedoma permafrost deposits in Siberian lowlands[17]. These cycles occur when abrupt permafrost thaw forms lakes that can drain over time, allowing sediments and carbon to refreeze into permafrost, while elsewhere new thaw lakes form and repeat this cyclic process (Fig. 4a, c). Abrupt thaw in upland regions, where water does not generally pool and form lakes, often creates gullies and slump features that can erode permafrost carbon into streams, rivers and lakes (Fig. 4b, d). These thaw features can also be widespread but are not as well recognized as are thaw lakes; over 7,500 upland thaw features were mapped within a 1,700-square-kilometre foothill region of Alaskan tundra[49]. Studies such as these illustrate a widespread influence of abrupt thaw in both upland and lowland permafrost landscapes, even though they do not provide a chronology of change.

Climate change is expected to increase the initiation and expansion of abrupt thaw features, potentially changing the rate of this historic disturbance cycle[62–65]. Wetland expansion due to abrupt thaw has affected 10% of a peatland landscape in northwestern Canada since the 1970s, with the fastest expansion occurring in the past decade[66]. Landscape lake cover is also affected by abrupt thaw, with net change being the sum of both lake expansion and drainage. The area of small open-water features around Prudhoe Bay on the Alaskan tundra has doubled since 1990 (ref. 67). In northwestern Alaska, lake initiation has increased since 1950, while lake expansion rates remained steady[68]. In general, landscape lake cover is currently believed to be stable or increasing within the continuous permafrost zone, whereas there is a tendency for lake drainage and vegetation infilling to dominate over lake expansion in the discontinuous permafrost zone[68–72].

Abrupt thaw influences carbon emissions to the atmosphere by exposing previously frozen carbon to microbial processes, and also by altering the hydrology that is critical for determining the balance of CO_2 and CH_4 emissions. Some of the highest CH_4 emissions in the permafrost region have been observed in lakes and wetlands formed through abrupt thaw[40,73]. At the same time, accumulation of new carbon under anaerobic conditions in peat[74] and in lake sediments[18] can be greater than permafrost carbon losses, at least in some ecosystems. In this way, anaerobic environments replace freezing temperatures as a mechanism for soil carbon stabilization, keeping greenhouse gas emissions lower than they would otherwise be[75]. In contrast, abrupt thaw processes in other landscapes clearly accelerate carbon loss. Drained lakes and lowered water tables will expose previously waterlogged carbon to microbial decomposition in aerobic conditions with relatively higher rates of carbon emissions. Also, lateral movement of permafrost carbon by leaching or erosion into lakes, rivers and the ocean[76–78] can increase loss, as carbon may be more readily mineralized through microbial and photochemical processes after mobilization[79,80]. How carbon cycling at the landscape scale will change under a warming climate will depend critically on how much of the landscape becomes wetter or drier, a question difficult to answer. It is clear that abrupt thaw is an important mechanism of rapid permafrost degradation, with widespread but varying influences on hydrology and carbon cycling. Yet abrupt thaw is not included in large-scale models, suggesting that important landscape transformations are not currently being considered in forecasts of permafrost carbon–climate feedbacks. This is in part due to the fact that we do not know at this stage what the relative importance of abrupt to gradual thaw across the landscape is likely to be.

Subsea carbon emissions

A majority of the observations and all of the modelling to date has focused on potential emissions from permafrost carbon on land. This is in part because subsea permafrost is buffered from recent climate change by the overlying ocean, and because ocean incursion at the end of the Ice Age has already been thawing and potentially reducing the pool of permafrost carbon under the sea. However, aside from organic carbon stored in permafrost, the sea bed underlying Arctic shelves also accumulated fossil CH_4 stored either as free CH_4 gas or as clathrates (CH_4-ice lattices that are stable at pressures and temperatures found at depth in this region). Layers of permafrost may serve as a physical barrier to the release of this CH_4 gas from the sediment into the water column and eventually the atmosphere. These

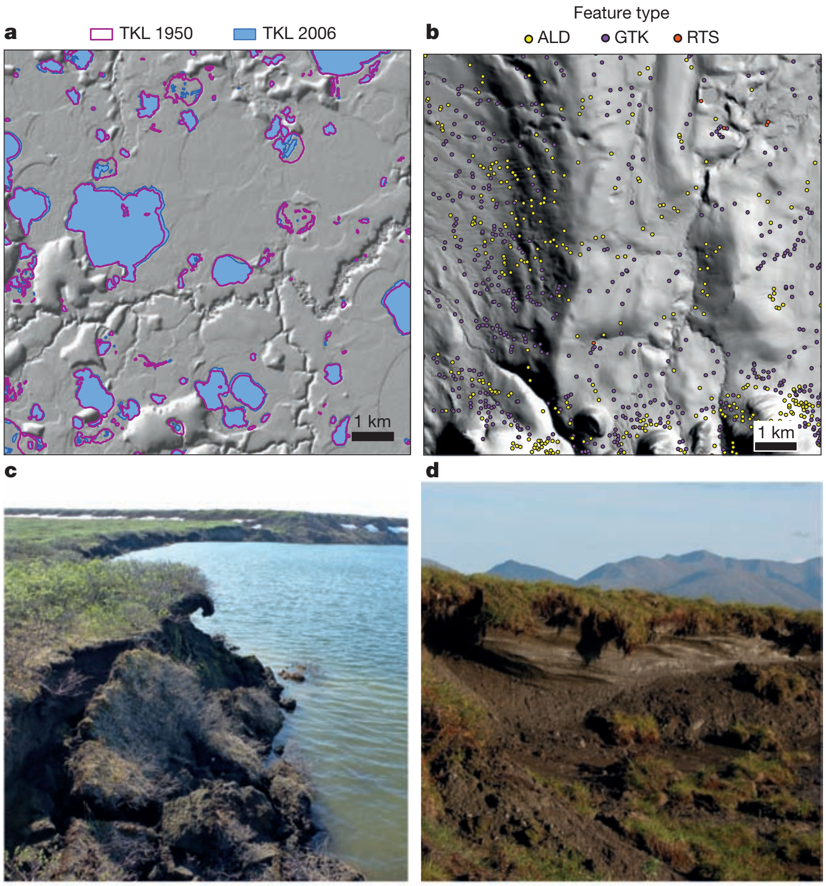


Figure 4 | Abundance of abrupt thaw features in lowland and upland settings in Alaska. Left panels (**a**, **c**) show thermokarst lake (TKL) abundance, expansion, and drainage on the Seward Peninsula, Northwest Alaska, between 1950 and 2006[68], with collapsing permafrost banks (photo credit G.G.). Right panels (**b**, **d**) show extensive distribution of ground collapse and erosion features (ALD, active layer detachment slide; RTS, retrogressive thaw slump; GTK, thermal erosion gullies) in upland tundra in a hill slope region in Northwest Alaska[61], and thawing icy soils in a retrogressive thaw slump (photo credit E.A.G.S.).

shallow shelves are also depositional areas for carbon from the erosion of coastal permafrost carbon and from inland permafrost carbon transported by Arctic rivers[81]. Together, these processes form ocean hotspots that are documented sources of high CH_4 emissions to the atmosphere[82,83], similar to hotspots formed in Arctic lakes on land[58]. New quantification has estimated that 17 Tg of CH_4 per year (where 1 Pg = 1,000 Tg) is emitted from the East Siberian Arctic Shelf after accounting for both diffusive and point-source bubble emissions[83]. Although this amount represents an increase from what was previously estimated for this region[27], this is probably because of improved observations of these emissions that may have been persistent over the thousands of years of land submergence. Climate warming, sea-ice decline, and increasing storminess have been linked to a 2.1 °C increase in bottom water (<10 m depth) temperature since the mid-1980s in this region[84]. Degradation of subsea permafrost from above by climate warming, and also from below by ongoing geothermal heat, will tend to increase new pathways between CH_4 storage areas deeper in the sediments and the sea floor[30]. But it is not known whether meaningful increases in CH_4 emissions via these processes could occur within this century, or whether they are more likely to manifest over a century or over millennia[84]. What is clear is that it would take thousands of years of CH_4 emissions at the current rate to release the same quantity of CH_4 (50 Pg) that was used in a modelled ten-year pulse to forecast tremendous global economic damage as a result of Arctic carbon release[8], making catastrophic impacts such as those appear highly unlikely[85–87].

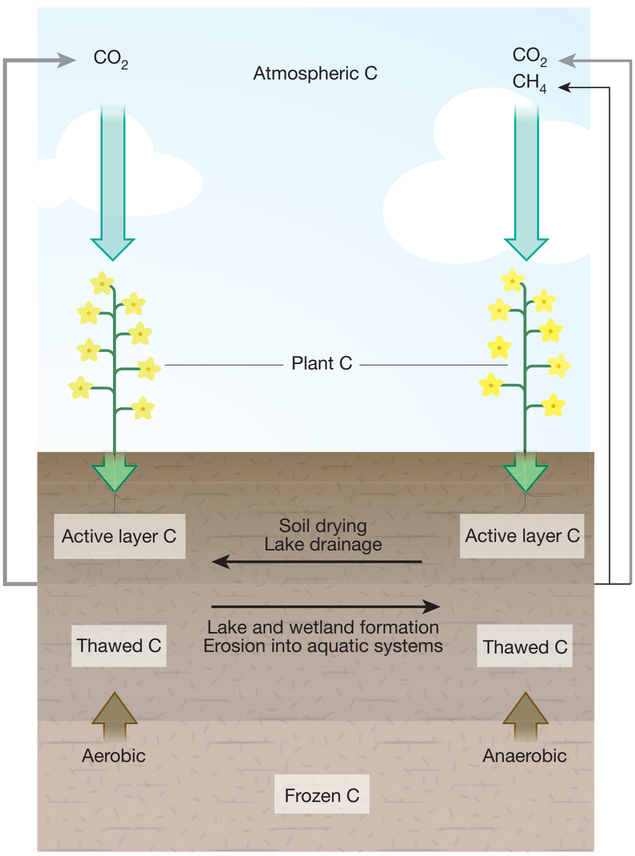
Permafrost and the global carbon cycle

Carbon pools in permafrost regions represent a large reservoir vulnerable to change in a warming climate. While some of this carbon will continue to persist in soils and sediments over the long term, our understanding that a substantial fraction of this pool is susceptible to microbial breakdown once thawed has been verified at the landscape scale (Box 1 and the Box 1 Figure). The exponential nature of microbial decomposition and CO_2 and CH_4 release over time means that the initial decades after thaw will be the most important for greenhouse gas release from any particular unit of thawed soil. Our expert judgement is that estimates made by independent approaches, including laboratory incubations, dynamic models, and expert assessment, seem to be converging on ~5%–15% of the terrestrial permafrost carbon pool being vulnerable to release in the form of greenhouse gases during this century under the current warming trajectory, with CO_2-carbon comprising the majority of the release. There is uncertainty, but the vulnerable fraction does not appear to be twice as high or half as much as 5%–15%, based on this analysis. Ten per cent of the known terrestrial permafrost carbon pool is equivalent to ~130–160 Pg carbon. That amount, if released primarily in the form of CO_2 at a constant rate over a century, would make it similar in magnitude to other historically important biospheric sources, such as land-use change (0.9 ± 0.5 Pg carbon per year; 2003–2012 average), but far less than fossil-fuel emissions[88] (9.7 ± 0.5 Pg carbon per year in 2012). Considering CH_4 as a fraction of permafrost carbon release would increase the warming impact of these emissions. At these rates, the observed and projected emissions of CO_2 and CH_4 from thawing permafrost are unlikely to occur at a speed that could cause abrupt climate change over a period of a few years to a decade[1,9]. A large pulse release of permafrost carbon on this timescale could cause climate change that would incur catastrophic costs to society[8], but there is little evidence from either current observations or model projections to support such a large and rapid pulse. Instead, permafrost carbon emissions are likely to occur over

BOX 1

Permafrost carbon feedback to climate change

As shown in the Box 1 Figure, carbon stored frozen in permafrost, once thawed, can enter ecosystems that have either predominantly aerobic (oxygen present) or predominantly anaerobic (oxygen limited) soil conditions. Across the permafrost region, there is a gradient of water saturation that ranges from mostly aerobic upland ecosystems to mostly anaerobic lowland lakes and wetlands. In aerobic soils, CO_2 is released by microbial decomposition of soil organic carbon, whereas both CO_2 and CH_4 are released from anaerobic soils and sediments. Microbial breakdown of soil organic carbon can happen in the surface active layer, which thaws each summer and refreezes in the winter, and in the subsurface as newly thawed carbon becomes available for decomposition after it has emerged from the perennially frozen pool. The decomposability of soil organic carbon varies across the landscape depending in part on the plant inputs as well as the soil environment, and also with depth in the soil profile. The landscape mosaic of water saturation is also affected by permafrost thaw. Gradual and abrupt thaw processes such as top-down thawing of permafrost (increasing the thickness of the active layer) and lake draining can expose more carbon to aerobic conditions. Alternatively, abrupt thaw processes can create wetter anaerobic conditions as the ground surface subsides, attracting local water. Carbon can also be mobilized by erosion or by leaching from upland soils into aquatic systems or sediments. Plant carbon uptake can be stored in increased plant biomass or deposited in the surface soils, which in part can offset losses from soils.



Box 1 Figure | Key features regulating the permafrost carbon feedback to climate from new, synthesized observations.

decades and centuries as the permafrost region warms, making climate change happen even faster than we project on the basis of emissions from human activities alone. Because of momentum in the system and the continued warming and thawing of permafrost, permafrost carbon emissions are likely not only during this century but also beyond. Although never likely to overshadow emissions from fossil fuel, each additional ton of carbon released from the permafrost region to the atmosphere will probably incur additional costs to society.

Next steps for model–data integration

The Earth system models analysed for the IPCC AR5[1] did not include permafrost carbon emissions, and there is a need for the next assessment to make substantive progress analysing this climate feedback. It is clear, even among models that are currently capable of simulating permafrost carbon emissions, that improvements are needed to the simulations of the physical and biological processes that control the dynamics of permafrost distribution and soil thermal regime[43,44,57]. The initial model projections we review here are based on a range of different model formulations, many of which are known to lack key structural features. Critical next steps that are being achieved by the research community include a permafrost carbon model intercomparison using standard driving variables to improve model formulations and conceptualization. Initial intercomparison results point towards several key structural features that should be implemented by models attempting to forecast permafrost carbon emissions. These include explicitly defining the vertical distribution of carbon in permafrost soils to account for the way atmospheric warming at the surface propagates through the soil, causing permafrost thaw and carbon decomposition at depth. Additionally, many large-scale models do not distinguish CH_4 versus CO_2 release and project only total carbon emissions. This partitioning depends on explicitly describing the interactions between permafrost thaw and surface hydrology and is critical to produce credible projections of the effect of permafrost carbon on climate. A first-order issue is whether the terrestrial landscape in the permafrost region, already interspersed with thaw lakes, wetlands and waterlogged soils, becomes wetter or drier in a warmer world[89]. Lastly, new modelling formulations for describing abrupt thaw are being developed. These are needed to understand how gradual warming from the surface, occurring across the entire landscape as currently modelled, compares to hotspots on the landscape where permafrost undergoes catastrophic ground collapse and rapid thaw. These issues go beyond temperature sensitivity alone and are at the forefront of current ecosystem model development and research.

Models are useful tools for making projections, but need to use observations more effectively for benchmarking and parameterization. Current models show a wide range of results when compared against benchmark data sets of permafrost soil temperatures[44], soil carbon stocks[90], and high-latitude carbon fluxes[91], emphasizing the high uncertainty in these projections. Now, new data sets on decomposability (reviewed here) are available and should be used to parameterize key aspects of model carbon feedbacks. The databases on decomposability however, remain two orders of magnitude smaller than surface (<1 m) carbon pool data sets. Increasing the number of laboratory incubations will help to constrain uncertainty regarding the potential for permafrost carbon to remain stable under different environmental conditions and will allow researchers to understand which controls over decomposition are most important for the slow turnover pools that comprise a large fraction of the total permafrost carbon pool. At the same time, further work is required to quantify the permafrost carbon pool itself better. Despite substantial recent progress, remote regions such as the Canadian High Arctic, central Siberia, and the subsea continental shelves remain poorly represented, with very few data points deeper than 1 m. Other data sets synthesizing field observations of CH_4 emissions and CO_2 exchange provide process-level understanding available for model validation as well[40,91–93]. Model–data fusion using these newly created databases from both laboratory and field observations is urgently needed to evaluate which models can credibly represent the permafrost region and thus help reduce the uncertainty in forecasting the permafrost carbon feedback.

High-latitude warming and the emission of permafrost carbon remains a likely global carbon cycle feedback to climate change. The sheer size of these frozen carbon pools and the rapid changes observed in the permafrost region warrant focused attention on these remote landscapes. The observations and modelling steps outlined here will help in forecasting future change. At the same time, it is imperative to continue developing effective observation networks, including remote sensing capability[94], to adequately quantify real-time CO_2 and CH_4 emissions from permafrost regions[95]. While increased permafrost carbon emissions in a warming climate are more likely to be gradual and sustained rather than abrupt and massive, such observation networks are needed to detect the potential emissions predicted here, and also to provide early warning of phenomena and potential surprises we do not yet fully appreciate or understand. The combination of robust observations with appropriate modelling tools for forecasting change is essential to properly evaluate permafrost carbon sources. The quantification of carbon sources in addition to those that are a direct result of human activity is necessary when developing and evaluating climate change mitigation policies.

Received 14 July 2014; accepted 12 February 2015.

1. IPCC in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) 1535 (Cambridge Univ. Press, 2013).
2. Brown, J. & Romanovsky, V. E. Report from the International Permafrost Association: state of permafrost in the first decade of the 21st century. *Permafr. Periglac. Process.* **19,** 255–260 (2008).
3. Romanovsky, V. E. *et al.* Thermal state of permafrost in Russia. *Permafr. Periglac. Process.* **21,** 136–155 (2010).
4. Romanovsky, V. E. *et al. Permafrost* (Arctic Report Card 2011) http://www.arctic.noaa.gov/report11/ (2013).
5. Zimov, S. A., Schuur, E. A. G. & Chapin, F. S. Permafrost and the global carbon budget. *Science* **312,** 1612–1613 (2006).
6. Tarnocai, C. *et al.* Soil organic carbon pools in the northern circumpolar permafrost region. *Glob. Biogeochem. Cycles* **23,** GB2023 (2009).
7. Schuur, E. A. G. *et al.* Vulnerability of permafrost carbon to climate change: implications for the global carbon cycle. *Bioscience* **58,** 701–714 (2008).
8. Whiteman, G., Hope, C. & Wadhams, P. Climate science: vast costs of Arctic change. *Nature* **499,** 401–403 (2013).
9. National Research Council. *Abrupt Impacts of Climate Change: Anticipating Surprises* (The National Academies Press, 2013).
10. Schädel, C. *et al.* Short communication on network related activities: research coordination network on the vulnerability of permafrost carbon. *Frozen Ground* **37,** 7 (2013).
11. Schuur, E. A. G. *et al.* Research coordination network on the vulnerability of permafrost carbon. *Frozen Ground* **35,** 6 (2011).
12. Hugelius, G. *et al.* Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps. *Biogeosciences* **11,** 6573–6593 (2014).
Revised and updated current state of knowledge on permafrost soil organic carbon stocks at circumpolar scales.
13. Zimov, S. A. *et al.* Permafrost carbon: stock and decomposability of a globally significant carbon pool. *Geophys. Res. Lett.* **33,** L20502 (2006).
14. Ping, C.-L. *et al.* High stocks of soil organic carbon in the North American Arctic region. *Nature Geosci.* **1,** 615–619 (2008).
15. Harden, J. W. *et al.* Field information links permafrost carbon to physical vulnerabilities of thawing. *Geophys. Res. Lett.* **39,** L15704 (2012).
Provides cumulative distributions of active layer thickness under current and future climates and estimates the amounts of newly thawed carbon and nitrogen.
16. Strauss, J. *et al.* The deep permafrost carbon pool of the yedoma region in Siberia and Alaska. *Geophys. Res. Lett.* **40,** 6165–6170 (2013).
Quantifies the organic carbon pool for yedoma deposits and thermokarst deposits in Siberia and Alaska.
17. Grosse, G. *et al.* Distribution of late Pleistocene ice-rich syngenetic permafrost of the Yedoma Suite in east and central Siberia, Russia. *US Geol. Surv. Open File Rep.* **2013–1078,** 1–37 (2013).
18. Walter Anthony, K. M. *et al.* A shift of thermokarst lakes from carbon sources to sinks during the Holocene epoch. *Nature* **511,** 452–456 (2014).
19. Mishra, U. *et al.* Empirical estimates to reduce modeling uncertainties of soil organic carbon in permafrost regions: a review of recent progress and remaining challenges. *Environ. Res. Lett.* **8,** 035020 (2013).
20. Hugelius, G. *et al.* The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions. *Earth Syst. Sci. Data* **5,** 3–13 (2013).
21. Jobbágy, E. G. & Jackson, R. B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* **10,** 423–436 (2000).
22. Schirrmeister, L. *et al.* Late Quaternary paleoenvironmental records from the western Lena Delta, Arctic Siberia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **299,** 175–196 (2011).
23. Kanevskiy, M., Shur, Y., Fortier, D., Jorgenson, M. T. & Stephani, E. Cryostratigraphy of late Pleistocene syngenetic permafrost (yedoma) in northern Alaska, Itkillik River exposure. *Quat. Res.* **75,** 584–596 (2011).
24. Schirrmeister, L. *et al.* Fossil organic matter characteristics in permafrost deposits of the northeast Siberian Arctic. *J. Geophys. Res. Biogeosci.* **116,** G00M02 (2011).
25. Hugelius, G. *et al.* High-resolution mapping of ecosystem carbon storage and potential effects of permafrost thaw in periglacial terrain, European Russian Arctic. *J. Geophys. Res. Biogeosci.* **116,** G03024 (2011).
26. Brosius, L. S. *et al.* Using the deuterium isotope composition of permafrost meltwater to constrain thermokarst lake contributions to atmospheric CH_4 during the last deglaciation. *J. Geophys. Res. Biogeosci.* **117,** G01022 (2012).
27. McGuire, A. D. *et al.* Sensitivity of the carbon cycle in the Arctic to climate change. *Ecol. Monogr.* **79,** 523–555 (2009).
28. Walter, K. M., Edwards, M. E., Grosse, G., Zimov, S. A. & Chapin, F. S. Thermokarst lakes as a source of atmospheric CH_4 during the last deglaciation. *Science* **318,** 633–636 (2007).
29. Romanovskii, N. N., Hubberten, H. W., Gavrilov, A. V., Eliseeva, A. A. & Tipenko, G. S. Offshore permafrost and gas hydrate stability zone on the shelf of East Siberian seas. *Geo-Mar. Lett.* **25,** 167–182 (2005).
30. Nicolsky, D. J. *et al.* Modeling sub-sea permafrost in the East Siberian Arctic Shelf: the Laptev Sea region. *J. Geophys. Res. Earth Surf.* **117,** F03028 (2012).
31. Dutta, K., Schuur, E. A. G., Neff, J. C. & Zimov, S. A. Potential carbon release from permafrost soils of Northeastern Siberia. *Glob. Change Biol.* **12,** 2336–2351 (2006).
32. Knoblauch, C., Beer, C., Sosnin, A., Wagner, D. & Pfeiffer, E.-M. Predicting long-term carbon mineralization and trace gas production from thawing permafrost of Northeast Siberia. *Glob. Change Biol.* **19,** 1160–1172 (2013).
33. Elberling, B. *et al.* Long-term CO_2 production following permafrost thaw. *Nature Clim. Change* **3,** 890–894 (2013).
34. Kirschbaum, M. U. F. Will changes in soil organic carbon act as a positive or negative feedback on global warming? *Biogeochemistry* **48,** 21–51 (2000).
35. Kätterer, T., Reichstein, M., Andren, O. & Lomander, A. Temperature dependence of organic matter decomposition: a critical review using literature data analyzed with different models. *Biol. Fertil. Soils* **27,** 258–262 (1998).
36. Schädel, C. *et al.* Circumpolar assessment of permafrost C quality and its vulnerability over time using long-term incubation data. *Glob. Change Biol.* **20,** 641–652 (2014).
Synthesizes the decomposability of permafrost organic matter using incubation data and calculates potential carbon loss for high-latitude soils.
37. Treat, C. *et al.* A pan-Arctic synthesis of CH_4 and CO_2 production from anoxic soil incubations. *Glob. Change Biol.* doi:10.1111/gcb.12875 (in the press).
38. Myhre, G. *et al.* in *Climate Change 2013: The Physical Science Basis. Contributions of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) 659–740 (Cambridge Univ. Press, 2013).
39. Verville, J. H., Hobbie, S. E., Iii, F. S. C. & Hooper, D. U. Response of tundra CH_4 and CO_2 flux to manipulation of temperature and vegetation. *Biogeochemistry* **41,** 215–235 (1998).
40. Olefeldt, D., Turetsky, M. R., Crill, P. M. & McGuire, A. D. Environmental and physical controls on northern terrestrial methane emissions across permafrost zones. *Glob. Change Biol.* **19,** 589–603 (2013).
Synthesis of data on growing-season CH_4 emissions from terrestrial ecosystems across permafrost zones.
41. Treat, C. C. *et al.* Temperature and peat type control CO_2 and CH_4 production in Alaskan permafrost peats. *Glob. Change Biol.* **20,** 2674–2686 (2014).
42. Ström, L., Tagesson, T., Mastepanov, M. & Christensen, T. R. Presence of *Eriophorum scheuchzeri* enhances substrate availability and methane emission in an Arctic wetland. *Soil Biol. Biochem.* **45,** 61–70 (2012).
43. Lawrence, D. M., Slater, A. G., Romanovsky, V. E. & Nicolsky, D. J. Sensitivity of a model projection of near-surface permafrost degradation to soil column depth and representation of soil organic matter. *J. Geophys. Res. Earth Surf.* **113,** F02011 (2008).
44. Koven, C. D., Riley, W. J. & Stern, A. Analysis of permafrost thermal dynamics and response to climate change in the CMIP5 Earth system models. *J. Clim.* **26,** 1877–1900 (2013).
Analysis of Earth system models projections of permafrost change in response to climate change scenarios.
45. Koven, C. D. *et al.* Permafrost carbon-climate feedbacks accelerate global warming. *Proc. Natl Acad. Sci. USA* **108,** 14769–14774 (2011).
46. Schaefer, K., Zhang, T., Bruhwiler, L. & Barrett, A. P. Amount and timing of permafrost carbon release in response to climate warming. *Tellus B* **63,** 165–180 (2011).
47. Schneider von Deimling, T. *et al.* Estimating the near-surface permafrost-carbon feedback on global warming. *Biogeosciences* **9,** 649–665 (2012).
48. MacDougall, A. H., Avis, C. A. & Weaver, A. J. Significant contribution to climate warming from the permafrost carbon feedback. *Nature Geosci.* **5,** 719–721 (2012).
49. Burke, E. J., Jones, C. D. & Koven, C. D. Estimating the permafrost-carbon climate response in the CMIP5 climate models using a simplified approach. *J. Clim.* **26,** 4897–4909 (2013).
50. Schaphoff, S. *et al.* Contribution of permafrost soils to the global carbon budget. *Environ. Res. Lett.* **8,** 014026 (2013).
51. Burke, E. J., Hartley, I. P. & Jones, C. D. Uncertainties in the global temperature change caused by carbon release from permafrost thawing. *Cryosphere* **6,** 1063–1076 (2012).
52. Zhuang, Q. *et al.* CO_2 and CH_4 exchanges between land ecosystems and the atmosphere in northern high latitudes over the 21st century. *Geophys. Res. Lett.* **33,** L17403 (2006).

53. Schuur, E. A. G., Abbott, B., &. the Permafrost Carbon Network. Climate change: high risk of permafrost thaw. *Nature* **480,** 32–33 (2011).
54. Schuur, E. A. G. *et al.* Expert assessment of vulnerability of permafrost carbon to climate change. *Clim. Change* **119,** 359–374 (2013).
State of knowledge on changes in permafrost distribution and soil organic carbon stocks in response to climate warming based on expert survey.
55. Schaefer, K., Lantuit, H., Romanovsky, V. E., Schuur, E. A. G. & Witt, R. The impact of the permafrost carbon feedback on global climate. *Environ. Res. Lett.* **9,** 085003 (2014).
56. Lawrence, D. M., Slater, A. G. & Swenson, S. C. Simulation of present-day and future permafrost and seasonally frozen ground conditions in CCSM4. *J. Clim.* **25,** 2207–2225 (2012).
57. Slater, A. G. & Lawrence, D. M. Diagnosing present and future permafrost from climate models. *J. Clim.* **26,** 5608–5623 (2013).
58. Shaver, G. R. *et al.* Global warming and terrestrial ecosystems: a conceptual framework for analysis. *Bioscience* **50,** 871–882 (2000).
59. Sistla, S. A. *et al.* Long-term warming restructures Arctic tundra without changing net soil carbon storage. *Nature* **497,** 615–618 (2013).
60. Qian, H., Joseph, R. & Zeng, N. Enhanced terrestrial carbon uptake in the northern high latitudes in the 21st century from the Coupled Carbon Cycle Climate Model Intercomparison Project model projections. *Glob. Change Biol.* **16,** 641–656 (2010).
61. Krieger, K. E. *The topographic form and evolution of thermal erosion features: A first analysis using airborne and ground-based LiDAR in Arctic Alaska*. MSc thesis, Idaho State Univ. (2012).
62. Jorgenson, M. T., Shur, Y. L. & Pullman, E. R. Abrupt increase in permafrost degradation in Arctic Alaska. *Geophys. Res. Lett.* **33,** L02503 (2006).
63. Christensen, T. R. *et al.* Thawing sub-arctic permafrost: effects on vegetation and methane emissions. *Geophys. Res. Lett.* **31,** L04501 (2004).
64. Johansson, T. *et al.* Decadal vegetation changes in a northern peatland, greenhouse gas fluxes and net radiative forcing. *Glob. Change Biol.* **12,** 2352–2369 (2006).
65. Osterkamp, T. E. Characteristics of the recent warming of permafrost in Alaska. *J. Geophys. Res. Earth Surf.* **112,** F02S02 (2007).
66. Baltzer, J. L., Veness, T., Chasmer, L. E., Sniderhan, A. E. & Quinton, W. L. Forests on thawing permafrost: fragmentation, edge effects, and net forest loss. *Glob. Change Biol.* **20,** 824–834 (2014).
67. Raynolds, M. K. *et al.* Cumulative geoecological effects of 62 years of infrastructure and climate change in ice-rich permafrost landscapes, Prudhoe Bay Oilfield, Alaska. *Glob. Change Biol.* **20,** 1211–1224 (2014).
68. Jones, B. M. *et al.* Modern thermokarst lake dynamics in the continuous permafrost zone, northern Seward Peninsula, Alaska. *J. Geophys. Res. Biogeosci.* **116,** G00M03 (2011).
69. Smith, L. C., Sheng, Y., MacDonald, G. M. & Hinzman, L. D. Disappearing Arctic lakes. *Science* **308,** 1429 (2005).
70. Riordan, B., Verbyla, D. & McGuire, A. D. Shrinking ponds in subarctic Alaska based on 1950–2002 remotely sensed images. *J. Geophys. Res. Biogeosci.* **111,** G04002 (2006).
71. Roach, J., Griffith, B., Verbyla, D. & Jones, J. Mechanisms influencing changes in lake area in Alaskan boreal forest. *Glob. Change Biol.* **17,** 2567–2583 (2011).
72. Sannel, A. B. K. & Kuhry, P. Warming-induced destabilization of peat plateau/thermokarst lake complexes. *J. Geophys. Res. Biogeosci.* **116,** G03035 (2011).
73. Walter, K. M., Zimov, S. A., Chanton, J. P., Verbyla, D. & Chapin, F. S. III. Methane bubbling from Siberian thaw lakes as a positive feedback to climate warming. *Nature* **443,** 71–75 (2006).
74. Jones, M. C., Grosse, G., Jones, B. M. & Walter Anthony, K. Peat accumulation in drained thermokarst lake basins in continuous, ice-rich permafrost, northern Seward Peninsula, Alaska. *J. Geophys. Res.* **117,** G00M07 (2012).
75. Zona, D. *et al.* Increased CO_2 loss from vegetated drained lake tundra ecosystems due to flooding. *Glob. Biogeochem. Cycles* **26,** GB2004 (2012).
76. Olefeldt, D. & Roulet, N. T. Permafrost conditions in peatlands regulate magnitude, timing, and chemical composition of catchment dissolved organic carbon export. *Glob. Change Biol.* **20,** 3122–3136 (2014).
77. Feng, X. *et al.* Differential mobilization of terrestrial carbon pools in Eurasian Arctic river basins. *Proc. Natl Acad. Sci. USA* **110,** 14168–14173 (2013).
78. Vonk, J. E. & Gustafsson, O. Permafrost-carbon complexities. *Nature Geosci.* **6,** 675–676 (2013).
79. Vonk, J. E. *et al.* High biolability of ancient permafrost carbon upon thaw. *Geophys. Res. Lett.* **40,** 2689–2693 (2013).
80. Cory, R. M., Crump, B. C., Dobkowski, J. A. & Kling, G. W. Surface exposure to sunlight stimulates CO_2 release from permafrost soil carbon in the Arctic. *Proc. Natl Acad. Sci. USA* **110,** 3429–3434 (2013).
81. Vonk, J. E. *et al.* Activation of old carbon by erosion of coastal and subsea permafrost in Arctic Siberia. *Nature* **489,** 137–140 (2012).
82. Shakhova, N. *et al.* Geochemical and geophysical evidence of methane release over the East Siberian Arctic Shelf. *J. Geophys. Res. Oceans* **115,** C08007 (2010).
83. Shakhova, N. *et al.* Ebullition and storm-induced methane release from the East Siberian Arctic Shelf. *Nature Geosci.* **7,** 64–70 (2014).
Quantitative assessment of bubble-induced CH_4 emissions resulting from subsea permafrost degradation in the coastal area.
84. Dmitrenko, I. A. *et al.* Recent changes in shelf hydrography in the Siberian Arctic: Potential for subsea permafrost instability. *J. Geophys. Res. Oceans* **116,** C10027 (2011).
85. Parmentier, F.-J. W. *et al.* The impact of lower sea-ice extent on Arctic greenhouse-gas exchange. *Nature Clim. Change* **3,** 195–202 (2013).
86. Notz, D., Brovkin, V. & Heimann, M. Arctic: uncertainties in methane link. *Nature* **500,** 529–529 (2013).
87. Parmentier, F.-J. W. & Christensen, T. R. Arctic: speed of methane release. *Nature* **500,** 529–529 (2013).
88. Le Quéré, C. *et al.* Global carbon budget 2013. *Earth Syst. Sci. Data* **6,** 235–263 (2014).
89. Avis, C. A., Weaver, A. J. & Meissner, K. J. Reduction in areal extent of high-latitude wetlands in response to permafrost thaw. *Nature Geosci.* **4,** 444–448 (2011).
90. Todd-Brown, K. E. O. *et al.* Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences* **10,** 1717–1736 (2013).
91. McGuire, A. D. *et al.* An assessment of the carbon balance of Arctic tundra: comparisons among observations, process models, and atmospheric inversions. *Biogeosciences* **9,** 3185–3204 (2012).
92. Belshe, E. F., Schuur, E. A. G. & Bolker, B. M. Tundra ecosystems observed to be CO_2 sources due to differential amplification of the carbon cycle. *Ecol. Lett.* **16,** 1307–1315 (2013).
93. Ueyama, M. *et al.* Upscaling terrestrial carbon dioxide fluxes in Alaska with satellite remote sensing and support vector regression. *J. Geophys. Res. Biogeosci.* **118,** 1266–1281 (2013).
94. National Research Council. *Opportunities to Use Remote Sensing in Understanding Permafrost and Related Ecological Characteristics: Report of a Workshop* http://dels.nationalacademies.org/Report/Opportunities-Remote-Sensing/18711 (The National Academies Press, 2014).
95. Schaefer, K., Lantuit, H., Romanovsky, V. E. & Schuur, E. A. G. *Policy Implications of Warming Permafrost* (United Nations Environment Program, 2012).
96. Brown, J., Ferrians, O. J. J., Heginbottom, J. A. & Melnikov, E. S. *Circum-Arctic Map of Permafrost and Ground-Ice Conditions*. Version 2, http://nsidc.org/data/docs/fgdc/ggd318_map_circumarctic/index.html (National Snow and Ice Data Center, 2002).
97. IPCC in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change* (eds Solomon, S. D. *et al.*) (Cambridge Univ. Press, 2007).

Acknowledgements Initial funding was provided by the National Science Foundation Vulnerability of Permafrost Carbon Research Coordination Network Grant number 955713, with continued support from the National Science Foundation Research, Synthesis, and Knowledge Transfer in a Changing Arctic: Science Support for the Study of Environmental Arctic Change Grant number 1331083. Author contributions were also supported by grants to individuals: Department of Energy Office of Science, Office of Biological and Environmental Sciences Division Terrestrial Ecosystem Sciences program (DE-SC0006982) to E.A.G.S.; National Science Foundation Long Term Ecological Research Program (1026415) to A.D.M.; Department of Energy (DE-AC02-05CH11231, NGEE Arctic, BGC-Feedbacks SFA) to C.D.K.; Regional and Global Climate Modeling Program (RGCM) of the US Department of Energy's Office of Science (BER) Cooperative Agreement (DE-FC02-97ER62402) to D.M.L.; European Research Commission (338335) to G.G.; The Netherlands Organization for Scientific Research (863.12.004) to J.E.V.; National Science Foundation Polar Programs (1312402) to S.M.N.; National Science Foundation Polar Programs (856864 and 1304271) to V.E.R.; National Oceanic and Atmospheric Administration (NA09OAR4310063) and National Aeronautics and Space Agency (NNX10AR63G) to K.S.; Nordforsk (DEFROST; 23001), EU FP7 (PAGE21; 282700) and FORMAS (Bolin Climate Research Centre; 214-2006-1749) to G.H. and P.K.; Department of Energy Biological and Environmental Research (3ERKP818) to D.J.H.; National Science Foundation, Division of Environmental Biology (724514, 830997) to M.R.T. and A.D.M.; U.S. Geological Survey Climate and Land Use Program to J.W.H. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government.

Author Contributions This manuscript arose from the collective effort of the Permafrost Carbon Network (http://www.permafrostcarbon.org); all authors are working group leaders within the network. E.A.G.S. and A.D.M. wrote the initial draft, with additional contributions from all authors. C.S. provided assistance with final editing and submission of the manuscript, and helped to organise the Permafrost Carbon Network activities that made this possible. Figure 1 was prepared by G.H., Fig. 2 by C.S., Fig. 3 by K.S., Fig. 4 by G.G. and the Box 1 Figure by E.A.G.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.A.G.S. (ted.schuur@nau.edu).

ARTICLE

doi:10.1038/nature14251

Branch-specific dendritic Ca^{2+} spikes cause persistent synaptic plasticity

Joseph Cichon[1] & Wen-Biao Gan[1]

The brain has an extraordinary capacity for memory storage, but how it stores new information without disrupting previously acquired memories remains unknown. Here we show that different motor learning tasks induce dendritic Ca^{2+} spikes on different apical tuft branches of individual layer V pyramidal neurons in the mouse motor cortex. These task-related, branch-specific Ca^{2+} spikes cause long-lasting potentiation of postsynaptic dendritic spines active at the time of spike generation. When somatostatin-expressing interneurons are inactivated, different motor tasks frequently induce Ca^{2+} spikes on the same branches. On those branches, spines potentiated during one task are depotentiated when they are active seconds before Ca^{2+} spikes induced by another task. Concomitantly, increased neuronal activity and performance improvement after learning one task are disrupted when another task is learned. These findings indicate that dendritic-branch-specific generation of Ca^{2+} spikes is crucial for establishing long-lasting synaptic plasticity, thereby facilitating information storage associated with different learning experiences.

One remarkable feature of the brain is to encode and store new information continuously without disrupting previously acquired memories. It is believed that experience-dependent changes in synaptic strength are crucial for information storage in the brain[1,2]. However, it remains unclear whether and how synaptic plasticity induced by past experiences are maintained in the face of new experiences[1–3]. To address this question, we examined the generation of dendritic Ca^{2+} spikes and their effect on synaptic plasticity in the primary motor cortex of mice performing different motor learning tasks.

Dendritic Ca^{2+} spikes trigger large Ca^{2+} influx into dendrites[4–8], and have been linked to activity-dependent increases or decreases of synaptic strength in brain slices[9–14]. Recent studies have shown that NMDA (*N*-methyl-D-aspartate)-receptor-dependent dendritic Ca^{2+} spikes are generated in several brain regions and are involved in the integration and amplification of synaptic inputs[15–19]. The function of Ca^{2+} spikes in regulating experience-dependent synaptic plasticity in the living brain remains undetermined. Here we show that different motor tasks induce dendritic Ca^{2+} spikes on different apical tuft branches of individual layer V (L5) pyramidal neurons in the mouse motor cortex. This spatial segregation of Ca^{2+} spikes is crucial for the induction and maintenance of synaptic potentiation related to different learning tasks. These findings underscore the important role of branch-specific dendritic Ca^{2+} spikes (BSDCS) in storing new information without disrupting existing memories in the brain.

Motor learning induces BSDCS

To investigate the potential role of dendritic Ca^{2+} spikes in learning-dependent synaptic plasticity, we first examined Ca^{2+} spike generation in the motor cortex of mice using a treadmill training paradigm[20]. In this model, head-restrained mice ran on a treadmill under a two-photon microscope and learned to change their gait patterns within 1 hour[20,21] (Fig. 1a–c and Extended Data Fig. 1). To detect Ca^{2+} spikes, we performed two-photon Ca^{2+} imaging of apical tuft dendrites of L5 pyramidal neurons expressing genetically encoded Ca^{2+} indicators (GCaMPs) (Fig. 1d and Extended Data Fig. 2a–c). We observed a significant increase (>700%) in the number of Ca^{2+} transients occurring across long stretches of dendrites in mice running on the treadmill than in quiet awake state (Fig. 1d–f and Extended Data Fig. 2c–h). As expected[22], dendritic Ca^{2+} transients detected with GCaMP6s exhibited a significantly larger $\Delta F/F_0$ but slower kinetics than with other GCaMPs (Fig. 1e and Extended Data Fig. 2d–f). The observed dendritic Ca^{2+} transients are reminiscent of NMDA-receptor-dependent Ca^{2+} spikes reported in other brain regions[15–19] in that they typically lasted hundreds of milliseconds with comparable $\Delta F/F_0$ across long dendritic segments (>30 μm) (Fig. 1d, e and Extended Data Fig. 2c–g). Moreover, the frequency and peak amplitude of these dendritic Ca^{2+} transients were significantly reduced in the presence of the NMDA receptor antagonist MK801 (Fig. 1g, h and Extended Data Fig. 2h, i). Together, these observations suggest that treadmill motor learning induces NMDA-receptor-dependent Ca^{2+} spikes on apical tuft branches of L5 pyramidal neurons in the motor cortex.

When mice underwent forward running (FWR) and then backward running (BWR) (five 30-s trials for each direction), of the tuft branches that spiked and were located within 100 μm below the pial surface, ~95% exhibited Ca^{2+} spikes during either FWR or BWR, while only ~5% showed Ca^{2+} spikes during both running tasks (Fig. 1i). Furthermore, in mice trained to run in four directions, only ~10% of those tuft branches exhibited Ca^{2+} spikes in response to two or more tasks (Fig. 1i). Thus, different running tasks induce Ca^{2+} spikes on different tuft branches with little overlap.

In contrast to non-overlapping Ca^{2+} spikes on distal tuft branches, we observed substantial overlap of Ca^{2+} activities in apical dendritic trunks (nexus, near the base of tuft branches) and L5 somata when mice were subjected to two or four direction running (Fig. 1a, j, l and Extended Data Fig. 3a–e). This observation suggests that different motor tasks induce Ca^{2+} spikes on separate tuft branches of the same L5 pyramidal neurons. Indeed, out of 33 pairs of sibling branches located within 100 μm below the pial surface, only 2 pairs showed FWR- and BWR-induced Ca^{2+} spikes on the same branches, whereas the remaining 31 pairs exhibited no such overlap (Fig. 1k, m and Extended Data Fig. 3h). As these higher-order tuft branches converge towards the nexus, a larger overlap of FWR-and BWR-induced Ca^{2+} spikes was observed on sibling branches located 100–200 μm from the pial surface (Fig. 1l, m and Extended Data Fig. 3f, g, i, j). At this cortical depth, ~16% of Ca^{2+}

[1]Skirball Institute, Department of Neuroscience and Physiology, New York University School of Medicine, New York, New York 10016, USA.

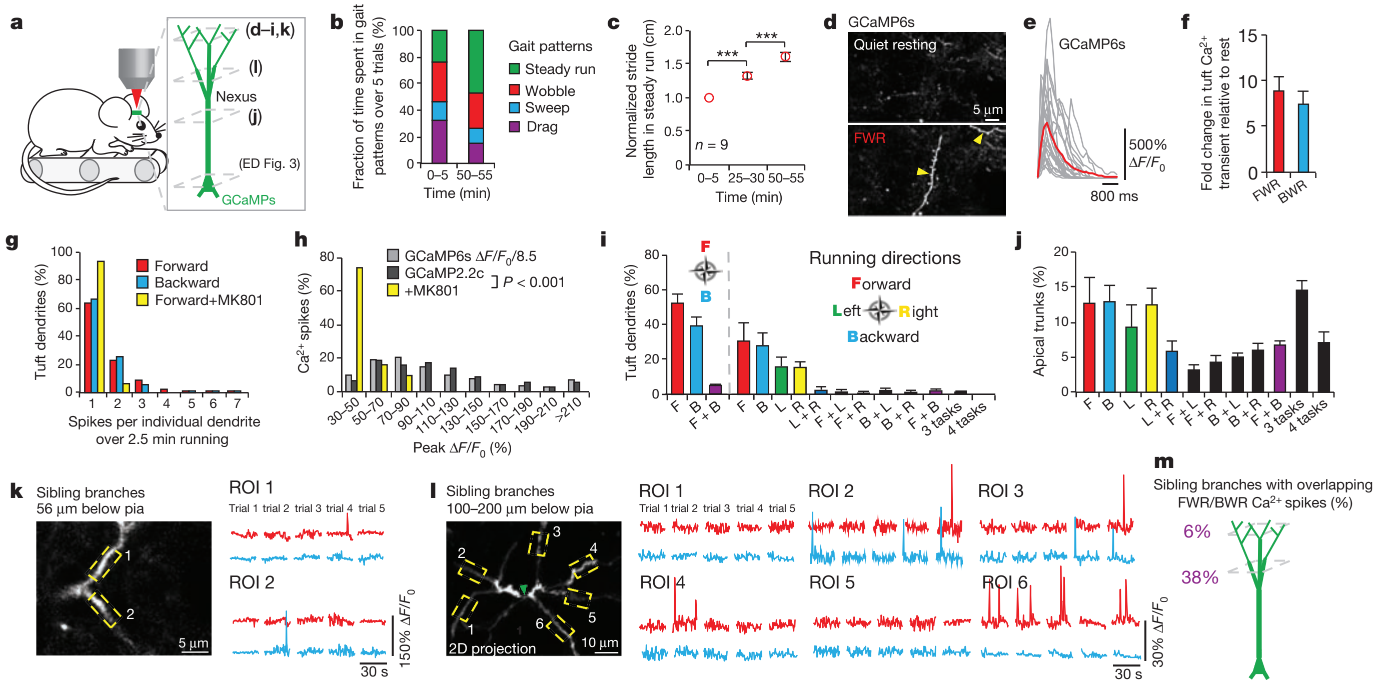


Figure 1 | Motor learning induces branch-specific Ca^{2+} spikes in apical tuft dendrites of L5 pyramidal neurons in motor cortex. **a**, Schematic of two-photon Ca^{2+} imaging at different cortical depths during treadmill training. **b**, Changes of forelimb gait patterns between early and late trials. **c**, Average stride length in steady run increased after training ($n = 9$ mice, paired t-test). **d**, Images of tuft dendrites expressing GCaMP6s during resting and FWR. Running-induced Ca^{2+} transients were visible over long dendritic segments (yellow arrowheads). **e**, Fast-scanning of individual Ca^{2+} transients (grey traces, $n = 24$) during FWR. Red trace represents the average. **f**, The number of Ca^{2+} transients increased >7-fold during FWR or BWR relative to resting ($P = 2.6 \times 10^{-7}$, paired t-test). **g**, Distribution of Ca^{2+} spike frequency on individual branches during FWR ($n = 321$), BWR ($n = 261$) and FWR with local MK801 ($n = 34$) over 2.5 min. **h**, Distribution of peak Ca^{2+} spike amplitudes detected with GCaMP6s and GCaMP2.2c during FWR with or without MK801 ($n = 141$, 213 and 31, respectively; $P < 0.0001$, Mann–Whitney test). $\Delta F/F_0$ values of Ca^{2+} spikes detected with GCaMP6s were ~8.5 times larger than with GCaMP2.2c. **i**, 95% of tuft branches that spiked were activated by only one task ($n = 616$ for two tasks; $n = 450$ for four tasks). **j**, 53% of apical trunks exhibited Ca^{2+} transients in response to several tasks ($n = 257$). **k**, Sibling branches exhibited non-overlapping FWR- and BWR-induced Ca^{2+} spikes. **l**, Two-dimensional projection of multiple sibling branches. Green arrowhead marks the trunk. Six regions of interests (ROIs) corresponding to different branches were analysed over five trials of FWR and BWR. **m**, The percentage of sibling branches with overlapping FWR/BWR-induced Ca^{2+} spikes at two cortical depths below the pia. Data are mean ± s.e.m. $^{***}P < 0.001$. See Methods for statistical details.

spikes (48 out of 294 spikes) occurred simultaneously in all branches (global) of the same neuron in response to FWR or BWR (Extended Data Fig. 3k). Notably, when tuft branches from an individual neuron were cut with a two-photon laser, running-induced Ca^{2+} activity in the trunk was significantly reduced (Extended Data Fig. 4). Together, these findings indicate that different motor tasks trigger Ca^{2+} spikes in largely non-overlapping distal apical tuft branches of the same L5 pyramidal neurons (Fig. 1m, Extended Data Fig. 3j and Supplementary Information). They also suggest that Ca^{2+} spikes generated in individual tuft branches propagate along dendrites and contribute to the activity in apical trunks and somata.

Spines active during BSDCS are potentiated

Dendritic Ca^{2+} spikes have been shown to trigger long-term synaptic potentiation or depression in brain slices[9–14]. To investigate the effect of Ca^{2+} spikes on synaptic plasticity *in vivo*, we examined Ca^{2+} activities of dendritic spines before and after Ca^{2+} spike generation. In mice subjected to FWR or BWR, Ca^{2+} transients in dendritic spines were readily detected with different GCaMP variants (Fig. 2a and Extended Data Fig. 5a–g). These running-induced spine Ca^{2+} transients sometimes coincided with the generation of Ca^{2+} spikes (Fig. 2a, b and Extended Data Fig. 5h–l). For spines active at the time of spike generation, the average ratio of GCaMP6s fluorescence intensity between spine heads and neighbouring shafts was 1.56 ± 0.13 (mean ± s.e.m.), significantly higher than that for neighbouring inactive spines (0.46 ± 0.02) (Extended Data Fig. 5k, l; $P < 0.001$). Notably, after Ca^{2+} spikes, these active spines exhibited a significant increase in the peak amplitude of Ca^{2+} transients (Fig. 2a–c and Extended Data Fig. 5h–j, m–o; 74.7 ± 7.3% for FWR ($n = 80$) and 52.7 ± 10.9% for BWR ($n = 18$); $P < 0.001$). By contrast, no significant potentiation of Ca^{2+} transients was observed in spines that were inactive at the time of spike generation or did not experience a Ca^{2+} spike (inactive spines: 1.8 ± 5.8%, $P = 0.76$, $n = 22$; no Ca^{2+} spike: 4.2 ± 2.9%, $P = 0.15$, $n = 35$). We also found that the increase in the peak amplitude of spine Ca^{2+} transients was significantly correlated with the peak amplitude of Ca^{2+} spikes and the ratio of fluorescence intensity between spine heads and neighbouring shafts at the time of spike generation (Fig. 2d, e and Supplementary Information). Moreover, local infusion of MK801 or calcium/calmodulin-dependent protein kinase II (CaMKII) inhibitors into the motor cortex abolished the potentiation of Ca^{2+} transients in spines active during FWR (Fig. 2c and Extended Data Fig. 6a). Unlike MK801, CaMKII inhibition did not significantly alter Ca^{2+} spike frequency or peak amplitude (Extended Data Fig. 6b–d). Together, these results show that Ca^{2+} spikes cause the potentiation of Ca^{2+} transients in synchronously active spines via a process requiring NMDA receptor activation and CaMKII signalling.

The potentiation of Ca^{2+} transients could be observed within 6.2 ± 0.7 s after a single Ca^{2+} spike induced by FWR ($n = 60$, Extended Data Fig. 5n). When spines were examined 20–40 min after Ca^{2+} spike-induced Ca^{2+} potentiation, most spines continued to exhibit Ca^{2+} potentiation (Fig. 2f; $P < 0.001$). We further examined dendritic spine size change over time in L5 pyramidal neurons expressing both GCaMP6s and a structural marker, tdTomato (Fig. 2g, h). Most spines that exhibited Ca^{2+} potentiation immediately after Ca^{2+} spikes also showed an increase in spine size within 40 min (Extended Data Fig. 5o). The average size of spines active during Ca^{2+} spikes increased over 40 min, but not over 2 min (Fig. 2g, h; 2 min: 2.1 ± 7.1%, $P = 0.77$, $n = 11$; 40 min: 32.3 ± 11.1%,

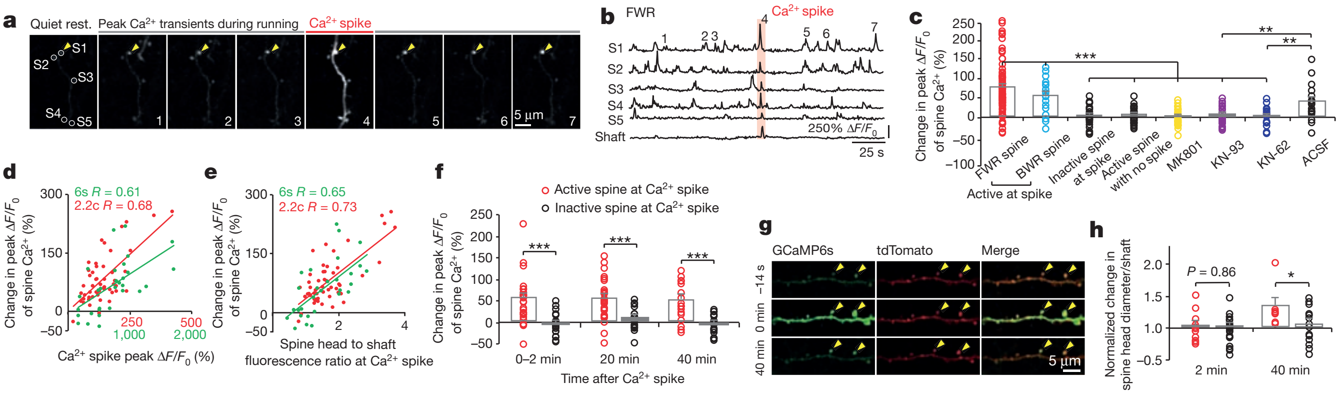


Figure 2 | Ca^{2+} spikes cause persistent potentiation of task-related dendritic spines. **a**, **b**, Time-lapse images and fluorescent traces of a representative apical tuft branch and spines expressing GCaMP6s. Running-induced Ca^{2+} transients in spines (arrowheads point to peak Ca^{2+} signals for spine 1) and a Ca^{2+} spike (marked by red bar). **c**, Spines active during spikes showed a significant increase in the peak amplitude of Ca^{2+} transients after Ca^{2+} spikes during FWR ($n = 80$) and BWR ($n = 18$), whereas spines not active during spikes ($P = 0.76$; $n = 22$ spines), or active spines with no encounter of spikes ($P = 0.15$; $n = 35$), showed no increase. Local application of MK801 ($n = 31$), KN-93 ($n = 20$) and KN-62 ($n = 18$), but not artificial cerebrospinal fluid (ACSF), blocked potentiation of spine Ca^{2+} transients ($P > 0.05$). **d**, **e**, Change in the peak amplitude of spine Ca^{2+} transients correlates with the peak amplitude of Ca^{2+} spike (**d**) and the fluorescence ratio between spine head and dendritic shaft (**e**) for spines active at the time of spike generation during FWR ($P < 0.001$ for GCaMP6s and GCaMP2.2c, Pearson correlation). **f**, Ca^{2+} spikes induced persistent potentiation of spine Ca^{2+} transients over 40 min ($n = 28$). **g**, **h**, Spines undergoing Ca^{2+} spikes-induced Ca^{2+} potentiation, not neighbouring inactive spines, also showed an increase in spine head size over 40 min ($n = 8$) but not over 2 min ($n = 11$). Data are mean ± s.e.m. *$P < 0.05$, **$P < 0.01$, *** $P < 0.001$, paired t-test. See Methods for statistical details.

$P < 0.05$; $n = 8$). There was no significant increase in the size of inactive spines over 2 or 40 min (2 min: 0.6 ± 4.5%, $P = 0.89$, $n = 23$; 40 min: 4.0 ± 7.6%, $P = 0.61$, $n = 14$). Because spine size strongly correlates with synaptic strength[23,24], these findings suggest that branch-specific Ca^{2+} spikes induced by motor running tasks cause long-lasting increase in the strength of synapses that are active at the time of spike generation.

Spines active before BSDCS are depotentiated

Previous *in vitro* studies have shown that Ca^{2+} spikes can cause either synaptic potentiation or depotentiation, depending on the time interval (typically less than 1 s) between synaptic activity and Ca^{2+} spike generation[25,26]. We therefore examined Ca^{2+} transients in spines that were active at different time intervals relative to Ca^{2+} spikes. In mice subjected to FWR, most spines (54%) exhibited activities coinciding with spike generation. In spines active asynchronously relative to Ca^{2+} spikes, the average time interval between the two events was 12.2 ± 1.6 s (Extended Data Fig. 7a, b). On the other hand, when animals were switched from FWR to BWR, many spine transients induced by FWR occurred seconds before the generation of Ca^{2+} spikes induced by BWR (Fig. 3a, b). Notably, FWR spines active within 5 s before Ca^{2+} spikes induced by BWR showed a significant reduction in the peak amplitude of Ca^{2+} transients when measured again in response to FWR (Fig. 3a–c; −38.6 ± 4.1%, $P < 0.001$; ratio of fluorescence intensity between spine heads and neighbouring shafts was 0.53 ± 0.03 during BWR-induced Ca^{2+} spikes; $n = 18$). There was no significant correlation between the degree of spine Ca^{2+} depotentiation and the peak amplitude of Ca^{2+} spikes (Fig. 3d; $R = -0.17$, $P = 0.35$). By contrast, FWR-activated spines did not show Ca^{2+} depotentiation if they were active >5 s before spike generation or did not encounter spikes (Fig. 3c; >5 s: −4.9 ± 3.4%, $P = 0.16$, $n = 31$; no spikes: 1.8 ± 5.0%, $P = 0.66$, $n = 31$; Extended Data Fig. 7c–f). Furthermore, local application of MK801 prevented Ca^{2+} spike-induced Ca^{2+} depotentiation in FWR-activated spines (Fig. 3c; −3.7 ± 3.2%, $P = 0.26$, $n = 34$). Together, these results indicate that Ca^{2+} spikes cause spine depotentiation if spines are active within 5 s before spike generation (Supplementary Information).

It is important to note that the percentage of spines active seconds before spike generation during FWR was much smaller than that of spines active at the time of spike generation (Extended Data Fig. 7a, b). As a result, Ca^{2+} spikes cause a net potentiation rather than depotentiation in spine Ca^{2+} transients during FWR. Moreover, because FWR and BWR induce Ca^{2+} spikes on different apical tuft branches, spines potentiated during FWR rarely encountered BWR-induced Ca^{2+} spikes to get depotentiated. Thus, the generation of Ca^{2+} spikes on different branches reduces the chance that synaptic changes induced by past experiences are disrupted by new experiences.

Persistent spine potentiation relies on BSDCS

The above findings suggest that the spatial segregation of dendritic Ca^{2+} spikes could be important for maintaining experience-dependent

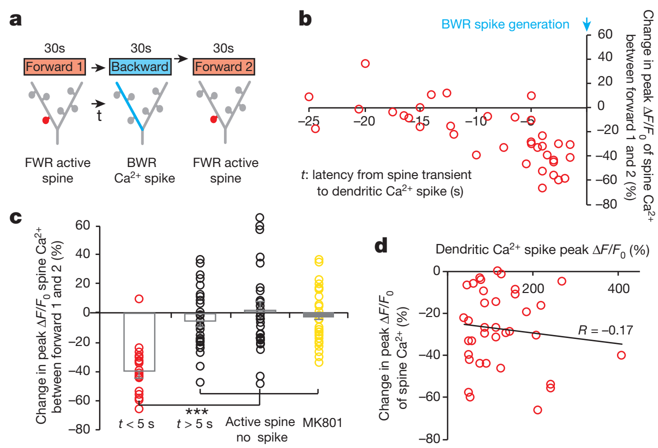


Figure 3 | Ca^{2+} spikes depotentiate spines active seconds before spike generation. **a**, A task-switching protocol for inducing spine activity and Ca^{2+} spike asynchronously. **b**, Changes in the peak amplitude of Ca^{2+} transients in FWR-activated spines versus the time interval between spine activity and a BWR-induced Ca^{2+} spike. **c**, Changes in the peak amplitude of spine Ca^{2+} transients under various conditions. Spines active within 5 s before the spike generation show a significant reduction in the peak amplitude of Ca^{2+} transients when measured again during FWR training ($n = 18$ spines), whereas spines active >5 s before the spike ($n = 31$ spines) or active spines without encountering spikes ($n = 31$ spines) showed no reduction. Local application of MK801 blocked the reduction. **d**, No correlation between the percentage change in the peak amplitude of spine Ca^{2+} transients and Ca^{2+} spike amplitude (Pearson correlation). Data are mean ± s.e.m. ***$P < 0.001$, paired t-test and Wilcoxon matched-pairs signed rank test. See Methods for statistical details.

synaptic plasticity. To investigate this further, we examined how the generation of branch-specific Ca^{2+} spikes is regulated and whether its disruption affects synaptic plasticity induced by different learning experiences.

Several lines of evidence suggest that the generation of Ca^{2+} spikes is regulated by cortical inhibition[27,28]. We found that local application of $GABA_A$ receptor antagonist bicuculline in layer I (L1) did not increase Ca^{2+} spike number but increased the overlap index of spikes induced by FWR and BWR (Extended Data Fig. 8a, b). We further examined the effects of deleting or silencing somatostatin (SST)-expressing cortical interneurons, which form synapses with apical tuft dendrites of L5 pyramidal neurons[29,30]. SST-interneuron deletion was accomplished with *SST-IRES-Cre/Rosa26-LSL-iDTR* mice in which diphtheria toxin treatment activated diphtheria toxin receptors (DTR)[31] and led to a specific loss ($69 \pm 4\%$) of SST neurons over 2 days (Fig. 4a and Extended Data Fig. 8c, d; before deletion: 877 cells; after deletion: 222 cells; $n = 5$ mice). SST-interneuron silencing was achieved with the designer receptors exclusively activated by designer drugs (DREADD) approach[32] in which activation of $G_{i/o}$-coupled human muscarinic M4 designer receptor by clozapine-*N*-oxide (CNO) induced a significant reduction in somatic Ca^{2+} activity of SST interneurons *in vivo* (Fig. 4b and Extended Data Fig. 8e; $P = 0.0001$; 27 somata; $n = 4$ mice). We found that either deletion or silencing of SST neurons caused loss of branch specificity of Ca^{2+} spikes without affecting the total number of Ca^{2+} spikes induced during running (Fig. 4c–e and Extended Data Fig. 8f; overlap index in control mice: <10%; overlap index in SST-deleted or silenced mice: ~40%). Thus, inactivating SST-interneurons disrupts branch-specific generation of Ca^{2+} spikes when mice perform different motor running tasks.

To test whether loss of branch-specificity of Ca^{2+} spikes affects the maintenance of synaptic plasticity, we first identified spines potentiated by Ca^{2+} spikes during FWR and then examined these potentiated spines during BWR in both SST-intact control and SST-deleted mice (Fig. 4f). Similar to control mice, the initial FWR session in SST-deleted mice induced potentiation ($82.6 \pm 8.8\%$) of Ca^{2+} transients in spines active at the time of spike generation (Extended Data Fig. 8g, h; $P < 0.001$; $n = 49$). Notably, during BWR, a larger fraction of FWR-potentiated spines underwent depotentiation in SST-deleted mice when compared to control mice (Fig. 4g; spines showing reduced $\Delta F/F_0$ after BWR training: 24 out of 34 (SST-deleted) and 8 out of 18 (SST-intact)). Detailed analyses show that many of these depotentiated spines were active within 5 s before Ca^{2+} spikes induced by BWR in SST-deleted mice (Fig. 4g, h). Owing to the loss of branch-specific generation of Ca^{2+} spikes, FWR-potentiated spines were depotentiated more often in SST-deleted mice than in control mice because they were more likely to be on the same dendrites in which BWR-induced Ca^{2+} spikes occurred.

In addition to being depotentiated, a fraction of FWR-potentiated spines were active during Ca^{2+} spikes induced by BWR and underwent further potentiation in SST-deleted mice (Fig. 4g, h). However, more spines were depotentiated than potentiated after BWR. Consequently, when BWR followed FWR, there was a net depotentiation of Ca^{2+} transients in FWR-potentiated spines in SST-deleted mice (Fig. 4i; $P < 0.05$; $n = 34$). In control mice, no such net depotentiation was observed ($P = 0.32$). Moreover, when the initial FWR was followed by a second FWR session, previously potentiated spines showed no net depotentiation in control mice and further net potentiation in SST-deleted mice (Fig. 4i; control: $P = 0.15$; SST-deleted: $P < 0.05$). Together, these findings indicate that the loss of branch-specific generation of Ca^{2+} spikes reduces synaptic potentiation induced by one task when a different task is learned.

Disrupting BSDCS causes learning interference

The functional consequences of disrupting branch-specific Ca^{2+} spikes and synaptic plasticity at apical tuft branches remain unclear. We found that local application of tetrodotoxin (TTX) to L1, but not L5, significantly reduced the number of tuft Ca^{2+} spikes as well as Ca^{2+} transients in the apical trunk nexus in mice running on the treadmill (Fig. 5a, b; $P < 0.001$). TTX application to L1 also reduced Ca^{2+} activity in L5 neuronal somata (Fig. 5c and Extended Data Fig. 9a–c). These observations are consistent with previous studies showing that synaptic inputs and Ca^{2+} spikes on tuft branches have a profound effect on the output of L5 pyramidal neurons by causing the depolarization in apical trunks and somata[33–35] (Extended Data Fig. 4 and Supplementary Information).

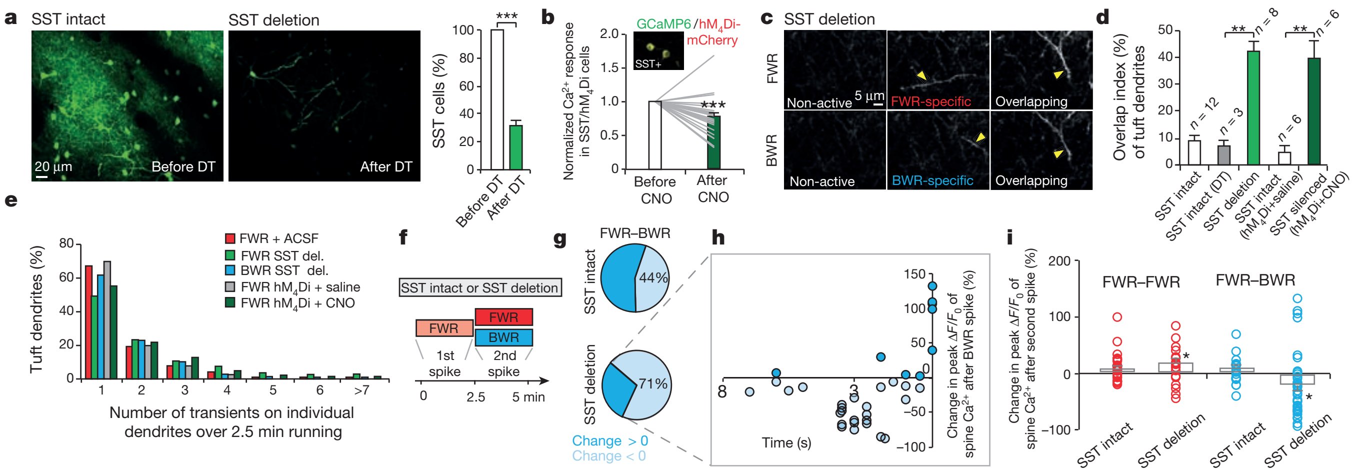


Figure 4 | Disrupting cortical inhibition alters branch-specific Ca^{2+} spikes and synaptic potentiation. **a**, Two-photon images of SST-interneurons labelled with green fluorescent protein (GFP) in motor cortex before and 2 days after diphtheria toxin (DT) treatment. Loss of SST interneurons occurred within 2 days after diphtheria toxin treatment. **b**, Somatic Ca^{2+} activity was reduced after CNO in SST interneurons expressing hM_4Di-mCherry/GCaMP6s (inset). **c**, Two-photon images of tuft dendrites exhibiting spikes during FWR and BWR after SST deletion. **d**, SST deletion or silencing significantly increased the percentage of tuft dendrites exhibiting Ca^{2+} spikes during both FWR and BWR (SST/DTR: $P = 0.0007$; SST/hM_4Di: $P = 0.002$). **e**, Inactivating SST interneurons did not affect the number of Ca^{2+} spikes generated during FWR as compared to control mice ($P > 0.05$). More than 200 Ca^{2+} spikes were measured per condition. del., deletion. **f**, Training protocol to test the effect of disrupting branch-specific Ca^{2+} spikes on potentiated spine Ca^{2+} transients in control and SST-deleted mice. **g**, Pie charts showing 44% and 71% of FWR-potentiated spines were depotentiated during BWR in control and SST-deleted mice, respectively. **h**, Spines potentiated during FWR were depotentiated when active <5 s before BWR-induced Ca^{2+} spikes in SST-deleted mice. **i**, Summary of the effect of FWR or BWR on previously potentiated spines in control and SST-deleted mice. Data are mean ± s.e.m. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, paired *t*-test (**a**, **d**, **i**) and Wilcoxon matched-pairs signed rank test (**b**, **i**). Statistical details in Methods.

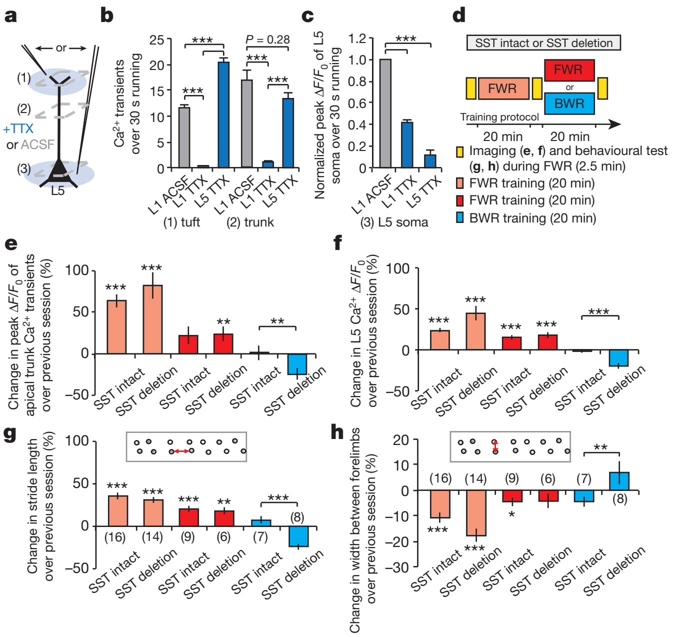


Figure 5 | Deletion of SST-interneurons impairs soma activities of L5 pyramidal neurons and performance improvement after learning. a–c, TTX applied to L1, but not L5, inhibited Ca^{2+} spike generation in the apical tuft and trunk. TTX to L1 also reduced Ca^{2+} transients in L5 somata during FWR. **d**, Experimental design to assess the effect of disrupting branch-specific Ca^{2+} spike on activities of apical trunks and somata, as well as behavioural performance in SST-deleted mice. **e**, **f**, In control and SST-deleted mice, Ca^{2+} activity at the trunk nexus (control: $n = 64$; deleted: $n = 61$) and L5 soma (control: $n = 113$; deleted: $n = 74$) increased after an initial and second sessions of FWR. BWR after the initial FWR significantly reduced Ca^{2+} activity at trunk (control: $n = 107$; deleted: $n = 77$) and L5 somata (control: $n = 63$; deleted: $n = 67$) in SST-deleted mice when compared to control mice. **g**, **h**, When FWR was followed by BWR, SST-deleted mice, but not control mice, displayed a reduction of the stride length and an increase in the width between forelimbs when tested in FWR. Number of mice per group is indicated. Data are mean ± s.e.m. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, Mann–Whitney U test (**b**–**f**) and paired t-test (**g**, **h**). See Methods for statistical details.

To test the functional effect of disrupting branch-specific Ca^{2+} spikes, we compared Ca^{2+} activities in apical trunks and L5 somata in control and SST-deleted mice subjected to FWR and/or BWR (Fig. 5d). In both control and SST-deleted mice, there was a substantial increase of Ca^{2+} activity in the trunk nexus and L5 somata during the first (0–20 min) and second (20–40 min) FWR training session (Fig. 5d–f). This increase of Ca^{2+} transients in trunks was blocked by local application of CaMKII inhibitors to L1 (Extended Data Fig. 9d, e). When the initial FWR was followed by BWR, there was no significant change in Ca^{2+} transients that were potentiated in the trunks and somata during the initial FWR session in control mice (Fig. 5d–f; change from 20–40 min: 1.0 ± 9.2% (trunk), $P = 0.12$; −1.0 ± 1.6% (L5 soma), $P = 0.10$). By contrast, in SST-deleted mice, there was a significant reduction in previously potentiated Ca^{2+} transients in both trunks and somata (Fig. 5d–f; change from 20–40 min: −24.2 ± 7.1% (trunk), $P < 0.001$; −19.8 ± 3.3% (L5 soma), $P < 0.001$). Together, these experiments indicate that in SST-deleted mice, a new learning task not only interferes with the maintenance of synaptic potentiation (Fig. 4) but also causes a reduction in the soma activities of L5 pyramidal neurons that are potentiated during a previously learned task.

When motor performance (stride length and width between forelimbs in steady FWR) was measured (Fig. 5d), control mice showed performance improvements when trained to FWR or BWR after the initial FWR (Fig. 5d, g, h). Similar to control mice, SST-deleted mice showed a progressive increase in the stride length and a decrease in the width between forelimbs when only running forward (Fig. 5d, g, h and Supplementary Information). However, when mice were subjected to BWR after FWR and their performances in FWR were tested again, SST-deleted mice exhibited a significant reduction in the stride length and an increase in the width between forelimbs when compared to control mice (Fig. 5d, g, h; $P < 0.001$ (length); $P < 0.01$ (width)). Additionally, when mice were subjected to motor learning on an accelerated rotarod[36], SST-deleted mice, but not the control mice, showed a significant reduction in FWR performance after BWR (Extended Data Fig. 10a). Together, these results suggest that disrupting branch-specific Ca^{2+} spikes and synaptic potentiation on apical tuft dendrites interferes with increased activities of L5 pyramidal neurons and behavioural improvement when multiple motor tasks are learned.

Discussion

How neural circuits encode and store vast amounts of information without significant interferences remains unclear. By performing Ca^{2+} imaging in the motor cortex of awake, behaving mice, we show that different motor learning experiences trigger NMDA-receptor-dependent Ca^{2+} spikes in non-overlapping apical tuft branches of L5 pyramidal neurons. This spatial segregation of Ca^{2+} spikes is crucial for the induction and maintenance of synaptic plasticity when several tasks are learned (Extended Data Fig. 10b). These findings indicate that dendritic-branch-specific generation of Ca^{2+} spikes is a fundamental mechanism used in the brain to store different information with little or no interference.

Our studies in the motor cortex are consistent with a view that individual dendritic branches serve as a basic unit for synaptic plasticity and information storage[37–40]. Recent studies have shown that NMDA-receptor-dependent Ca^{2+} spikes are involved in input integration and amplification in sensory cortices and hippocampus[15–19]. As in the motor cortex, NMDA-receptor-dependent Ca^{2+} spikes in these brain regions may also have an important role in experience-dependent synaptic plasticity. Unlike the motor cortex, however, Ca^{2+} spikes in the barrel cortex occur in all apical tuft branches of an individual L5 pyramidal neuron[15]. It would be interesting to investigate whether the lack of branch-specific generation of Ca^{2+} spikes in sensory cortices might compromise the maintenance of experience-dependent synaptic plasticity during sensory processing and perceptual learning. Furthermore, in addition to SST-expressing interneurons, the generation of branch-specific Ca^{2+} spikes and their effect on synaptic plasticity are probably regulated by other types of inhibitory neurons, ion channels and behavioural context[5,13,41–43]. Future studies are needed to investigate these issues to understand better how information is encoded and stored in the brain.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 June 2014; accepted 21 January 2015.
Published online 30 March 2015.

1. Martin, S. J., Grimwood, P. D. & Morris, R. G. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu. Rev. Neurosci.* **23,** 649–711 (2000).
2. Neves, G., Cooke, S. F. & Bliss, T. V. Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. *Nature Rev. Neurosci.* **9,** 65–75 (2008).
3. Abraham, W. C. & Robins, A. Memory retention–the synaptic stability versus plasticity dilemma. *Trends Neurosci.* **28,** 73–78 (2005).
4. Ross, W. N. & Werman, R. Mapping calcium transients in the dendrites of Purkinje cells from the guinea-pig cerebellum *in vitro*. *J. Physiol. (Lond.)* **389,** 319–336 (1987).
5. Jaffe, D. B. *et al.* The spread of Na^+ spikes determines the pattern of dendritic Ca^{2+} entry into hippocampal neurons. *Nature* **357,** 244–246 (1992).
6. Schiller, J., Schiller, Y., Stuart, G. & Sakmann, B. Calcium action potentials restricted to distal apical dendrites of rat neocortical pyramidal neurons. *J. Physiol. (Lond.)* **505,** 605–616 (1997).
7. Spruston, N. Pyramidal neurons: dendritic structure and synaptic integration. *Nature Rev. Neurosci.* **9,** 206–221 (2008).
8. Major, G., Larkum, M. E. & Schiller, J. Active properties of neocortical pyramidal neuron dendrites. *Annu. Rev. Neurosci.* **36,** 1–24 (2013).
9. Golding, N. L., Staff, N. P. & Spruston, N. Dendritic spikes as a mechanism for cooperative long-term potentiation. *Nature* **418,** 326–331 (2002).

10. Kampa, B. M., Letzkus, J. J. & Stuart, G. J. Requirement of dendritic calcium spikes for induction of spike-timing-dependent synaptic plasticity. *J. Physiol. (Lond.)* **574,** 283–290 (2006).
11. Remy, S. & Spruston, N. Dendritic spikes induce single-burst long-term potentiation. *Proc. Natl Acad. Sci. USA* **104,** 17192–17197 (2007).
12. Holthoff, K., Kovalchuk, Y., Yuste, R. & Konnerth, A. Single-shock LTD by local dendritic spikes in pyramidal neurons of mouse visual cortex. *J. Physiol. (Lond.)* **560,** 27–36 (2004).
13. Humeau, Y. & Luthi, A. Dendritic calcium spikes induce bi-directional synaptic plasticity in the lateral amygdala. *Neuropharmacology* **52,** 234–243 (2007).
14. Nevian, T. & Sakmann, B. Single spine Ca^{2+} signals evoked by coincident EPSPs and backpropagating action potentials in spiny stellate cells of layer 4 in the juvenile rat somatosensory barrel cortex. *J. Neurosci.* **24,** 1689–1699 (2004).
15. Xu, N. L. *et al.* Nonlinear dendritic integration of sensory and motor input during an active sensing task. *Nature* **492,** 247–251 (2012).
16. Lavzin, M., Rapoport, S., Polsky, A., Garion, L. & Schiller, J. Nonlinear dendritic processing determines angular tuning of barrel cortex neurons *in vivo*. *Nature* **490,** 397–401 (2012).
17. Smith, S. L., Smith, I. T., Branco, T. & Hausser, M. Dendritic spikes enhance stimulus selectivity in cortical neurons *in vivo*. *Nature* **503,** 115–120 (2013).
18. Grienberger, C., Chen, X. & Konnerth, A. NMDA receptor-dependent multidendrite Ca^{2+} spikes required for hippocampal burst firing *in vivo*. *Neuron* **81,** 1274–1281 (2014).
19. Palmer, L. M. *et al.* NMDA spikes enhance action potential generation during sensory input. *Nature Neurosci.* **17,** 383–390 (2014).
20. Yang, G. *et al.* Sleep promotes branch-specific formation of dendritic spines after learning. *Science* **344,** 1173–1178 (2014).
21. Marvin, J. S. *et al.* An optimized fluorescent probe for visualizing glutamate neurotransmission. *Nature Methods* **10,** 162–170 (2013).
22. Chen, T. W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499,** 295–300 (2013).
23. Harris, K. M. & Stevens, J. K. Dendritic spines of CA 1 pyramidal cells in the rat hippocampus: serial electron microscopy with reference to their biophysical characteristics. *J. Neurosci.* **9,** 2982–2997 (1989).
24. Matsuzaki, M. *et al.* Dendritic spine geometry is critical for AMPA receptor expression in hippocampal CA1 pyramidal neurons. *Nature Neurosci.* **4,** 1086–1092 (2001).
25. Sjöstrom, P. J. & Nelson, S. B. Spike timing, calcium signals and synaptic plasticity. *Curr. Opin. Neurobiol.* **12,** 305–314 (2002).
26. Lisman, J. & Spruston, N. Postsynaptic depolarization requirements for LTP and LTD: a critique of spike timing-dependent plasticity. *Nature Neurosci.* **8,** 839–841 (2005).
27. Murayama, M. *et al.* Dendritic encoding of sensory stimuli controlled by deep cortical interneurons. *Nature* **457,** 1137–1141 (2009).
28. Jiang, X., Wang, G., Lee, A. J., Stornetta, R. L. & Zhu, J. J. The organization of two new cortical interneuronal circuits. *Nature Neurosci.* **16,** 210–218 (2013).
29. Gentet, L. J. *et al.* Unique functional properties of somatostatin-expressing GABAergic neurons in mouse barrel cortex. *Nature Neurosci.* **15,** 607–612 (2012).
30. Chiu, C. Q. *et al.* Compartmentalization of GABAergic inhibition by dendritic spines. *Science* **340,** 759–762 (2013).
31. Saito, M. *et al.* Diphtheria toxin receptor-mediated conditional and targeted cell ablation in transgenic mice. *Nature Biotechnol.* **19,** 746–750 (2001).
32. Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S. & Roth, B. L. Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proc. Natl Acad. Sci. USA* **104,** 5163–5168 (2007).
33. Larkum, M. E., Nevian, T., Sandler, M., Polsky, A. & Schiller, J. Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. *Science* **325,** 756–760 (2009).
34. Harnett, M. T., Xu, N. L., Magee, J. C. & Williams, S. R. Potassium channels control the interaction between active dendritic integration compartments in layer 5 cortical pyramidal neurons. *Neuron* **79,** 516–529 (2013).
35. Yuste, R., Gutnick, M. J., Saar, D., Delaney, K. R. & Tank, D. W. Ca^{2+} accumulations in dendrites of neocortical pyramidal neurons: an apical band and evidence for two functional compartments. *Neuron* **13,** 23–43 (1994).
36. Buitrago, M. M., Schulz, J. B., Dichgans, J. & Luft, A. R. Short and long-term motor skill learning in an accelerated rotarod training paradigm. *Neurobiol. Learn. Mem.* **81,** 211–216 (2004).
37. Losonczy, A., Makara, J. K. & Magee, J. C. Compartmentalized dendritic plasticity and input feature storage in neurons. *Nature* **452,** 436–441 (2008).
38. Polsky, A., Mel, B. W. & Schiller, J. Computational subunits in thin dendrites of pyramidal cells. *Nature Neurosci.* **7,** 621–627 (2004).
39. Govindarajan, A., Israely, I., Huang, S. Y. & Tonegawa, S. The dendritic branch is the preferred integrative unit for protein synthesis-dependent LTP. *Neuron* **69,** 132–146 (2011).
40. Euler, T., Detwiler, P. B. & Denk, W. Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature* **418,** 845–852 (2002).
41. Lee, S., Kruglikov, I., Huang, Z. J., Fishell, G. & Rudy, B. A disinhibitory circuit mediates motor integration in the somatosensory cortex. *Nature Neurosci.* **16,** 1662–1670 (2013).
42. Magee, J., Hoffman, D., Colbert, C. & Johnston, D. Electrical and calcium signaling in dendrites of hippocampal pyramidal neurons. *Annu. Rev. Physiol.* **60,** 327–346 (1998).
43. Ahissar, E. *et al.* Dependence of cortical plasticity on correlated activity of single neurons and on behavioral context. *Science* **257,** 1412–1415 (1992).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by National Institutes of Health R01 NS047325 and P01 NS074972 to W.-B.G. We thank L. Looger, G. Chen and members of the Gan laboratory for their comments on the manuscript. We thank the Genetically-Encoded Neuronal Indicator and Effector (GENIE) Project and the Janelia Farm Research Campus of the Howard Hughes Medical Institute for sharing GCaMP6 constructs.

Author Contributions J.C. and W.-B.G. designed the experiments. J.C. performed experiments and analysed the data with the help from W.-B.G. J.C. and W.-B.G. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.-B.G. (gan@saturn.med.nyu.edu).

METHODS

Transgenic mice expressing GCaMP2.2c in L5 pyramidal neurons, *Thy1*-GCaMP2.2c (ref. 44), were received from G. Feng. Adeno-associated virus (AAV) experiments were conducted with C57/BL6 mice (Charles River Laboratory) and *Thy1*-Cre transgenic mice (FVB/N-Tg(Thy1-Cre)1Vln/J; the Jackson Laboratories). *SST-IRES-Cre* mice (*Sst$^{tm2.1(cre)Zjh}$*/J) and *Rosa26-stop-iDTR* mice were purchased from the Jackson Laboratories. Mice were housed in the Parasitology Central Animal Facility at New York University Medical Center. All of the 1–1.5-month-old male and female animals were used for the experiments in accordance with the Institutional Guidelines.

Treadmill training and behavioural analysis. A custom-built free-floating treadmill ($96 \times 56 \times 61$ cm dimensions) was used for motor training in this study[20,21]. This free-floating treadmill allows head-fixed mice to move their forelimbs freely to perform motor running tasks (forward, backward, left or right running). To minimize motion artefacts during imaging, the treadmill was constructed so that all the moving parts (motor, belt and drive shaft) have no contact with the microscope stage and the supporting air table. Animals were positioned on a custom-made head holder device that allowed the micro-metal bars (attached to the animal?s skull) to be mounted and for the base of the device to be positioned below the belt in contact with the microscope stage. During motor training, the treadmill motor (Dayton, Model 2L010) was driven by a DC power supply (Extech). At the onset of a trial, the motor was turned on and the belt speed gradually increased from 0 cm s^{-1} to 8 cm s^{-1} within ~3 s.

Each running trial lasted 30 s. After completion of each trial, the treadmill was turned off and the next trial started after 30 s rest period. A backward running was introduced to provide mice with new motor learning experience. In this model, animals were forced to run backward on the treadmill (speed increased gradually from 0 to 4 cm s^{-1} within ~3 s) for 30 s with the reversal of the treadmill motor. Left and right direction running was performed by shifting the custom-built head-holder device by 90° on the platform of the microscope, such that the mouse?s limbs were perpendicular to the treadmill?s belt. Turning on the treadmill caused the forelimbs of the head-restrained animal to move laterally. Mice completed five trials for each task, regardless of two task or four tasks running.

In Fig. 3, mice underwent a task-switching model in which mice were running forward, backward and then forward, all within a single trial (30 s for each direction, total running time 90 s). The task switching was accomplished quickly (within 1 s) by changing the polarity of the voltage applied to the motor. The belt speed was at 8 cm s^{-1} throughout the forward and backward running sessions.

To assay running gait patterns, mouse forelimbs were coated in ink (Speedball) before each running trial and white paper (45-cm long) was taped to the belt. Mice ran on white paper over five trials (30 s each) with 30 s inter-trial interval. The stride length and the width between forelimbs were analysed over the first 10 s of each 30 s trial. Footprints were analysed offline manually by three measures: (1) the fraction of time spent in gait pattern over five trials (four groups: drag, sweep, wobble and steady run); (2) the stride length between two ipsilateral forelimb footprints in steady run gait pattern; and (3) the width between two forelimbs. The stride length and the width between forelimbs were averaged over five trials per animal before and after two 20-min training epochs. In early forward running trials, mice exhibited various gait patterns including drag (35%), sweep (15%), wobble (30%) and steady run (20%), and with continued training, mice refined their gait patterns from drag (13%), sweep (12%) and wobble (29%) into steady run (46%) (Fig. 1b and Extended Data Fig. 1). We define drag as a gait pattern in which mice failed to execute steps, which resulted in a smearing of ink during the trial run. Sweep occurs as mice make distinct, yet lateral forelimb steps resulting ink marks perpendicular to the direction of running. Wobble occurs when mice make many disorganized, chaotic steps resulting in an ink pattern that appears highly variable in and lateral to the running direction. During steady run, mice make distinct steps resulting in spaced ink marks centred in the direction of the running trial. As shown in Fig. 1 and Extended Data Fig. 1, we observed changes in the stride length and the width between forelimbs when mice achieved steady run gait over forward training. Similar to forward running, backward running also elicited changes in gait patterns and the stride length over 40-min training (see Extended Data Fig. 1). The primary motor cortex is necessary for this task as local pharmacological inactivation with muscimol prevented the emergence of steady run gait patterning (Extended Data Fig. 1e).

Rotarod training and behavioural analysis. An EZRod system with a test chamber ($44.5 \times 14 \times 51$ cm, Accuscan Instruments) was used as another motor training model different from treadmill running[20]. Animals were placed on the motorized rod (30 mm in diameter) in the chamber. The rotation speed gradually increased from 0 to 100 r.p.m. over the course of 3 min. The time latency and rotation speed were recorded when the animal was unable to keep up with the increasing speed and fell from the rod. Rotarod training was performed in two training sessions each consisting of 20 trials. Rotarod performance was tested in five trials before the training as well as after the first and second training session. Performance was measured as the average speed animals achieved during the five trials at each time point. To test motor interference, backward running was introduced after the first forward running session. In this regime, animals were subjected to run backward on the rotating rod (speed increased gradually from 0 to 40 r.p.m. over 3 min) for 20 trials. After backward running, animals were tested for their forward running performance (Extended Data Fig. 10a).

Surgical preparation for imaging awake, head-restrained mice. Dendritic imaging was carried out in awake, head-restrained mice through a thinned-skull window. Surgery preparation for awake animal imaging includes attaching the head holder and creating a thinned-skull cranial window[45]. Specifically, mice were deeply anaesthetized with an intraperitoneal injection of ketamine (100 μg g^{-1}) and xylazine (10 μg g^{-1}). The mouse head was shaved and the skull surface was exposed with a midline scalp incision. The periosteum tissue over the skull surface was removed without damaging the temporal and occipital muscles.

A head holder composed of two parallel micro-metal bars was attached to the animal?s skull to help restrain the animal?s head and reduce motion-induced artefact during imaging. A small skull region (~0.2 mm in diameter) was located over the primary motor cortex based on stereotactic coordinates (at bregma and 1.2 mm lateral from the midline) and marked with a pencil[46]. A thin layer of cyanoacrylate-based glue was first applied to the top of the entire skull surface, and the head holder was then mounted on top of the skull with dental acrylic cement (Lang Dental Manufacturing Co.) such that the marked skull region was exposed between the two bars. Precaution was taken not to cover the marked region with dental acrylic cement.

After the dental cement was completely dry, the head holder was screwed to two metal cubes that were attached to a solid metal base, and a cranial window was created over the previously marked region. The procedures for preparing a thinned skull cranial window for two-photon imaging have been described in detail previously[45]. In brief, a high-speed drill was used to reduce carefully the skull thickness by approximately 50% under a dissecting microscope. The skull was immersed in artificial cerebrospinal fluid (ACSF) during drilling. Skull thinning was completed by carefully scraping the cranial surface with a microsurgical blade to ~20 μm in thickness. A high quality picture of the brain vasculature was taken with a CCD camera attached to a stereo dissecting microscope. The completed cranial window was covered with silicone elastomer (World Precision Instruments) and the animals were returned to their own cages to recover. For calcium imaging of L5 somata, a small piece of bone overlying the primary motor cortex was removed, and a round glass coverslip (approximately the same size as the bone being removed) was glued to the skull to reduce motion of the exposed brain.

Imaging experiments were started ~12–24 h after window implantation. After awakening, mice with head mounts were habituated for a few times (10 min each time) in the treadmill-adapted imaging apparatus to minimize potential stress effects of head restraining, motor training and imaging. Awake mice were head restrained in the imaging apparatus, which sits on top of a custom-built free-floating treadmill.

Two-photon calcium imaging of dendrites and somata of L5 neurons expressing GCaMPs. Genetically encoded calcium indicators including GCaMP6s, GCaMP5G, GCaMP3 and GCaMP2.2c were used for Ca^{2+} imaging of dendritic spikes of L5 pyramidal neurons in the primary motor cortex (a region ~200 μm in square and centred at ~0.5 mm anterior from bregma and ~1.2 mm lateral from the midline). GCaMPs were expressed with recombinant AAV under the human synapsin-1 (*SYN1*) promoter (AAV, serotype 2/1; $>2 \times 10^{13}$ genome copies per ml; produced by the University of Pennsylvania Gene Therapy Program Vector Core). A total of 0.1–0.2 μl of AAV viruses was diluted 10 times in ACSF and slowly injected (Picospritzer III; 15 p.s.i., 12 ms, 0.8 Hz) over 10–15 min into L5 (depth of 500–700 μm) of primary motor cortex using a glass microelectrode around the coordinate of 0.5 cm anterior and 1.5 cm lateral from bregma[46]. The ACSF dilution allowed better spread of viruses through L5 and sparse neuronal labelling. Sparse expression of GCaMP ensures that the contribution of fluorescent signals from the neuropil to the dendrites and somata is negligible. In addition to *SYN1*-driven GCaMPs, *Thy1*-Cre transgenic mice (FVB/N-Tg(Thy1-Cre)1Vln/J; Jackson stock number: 6143) were used in combination with Cre-dependent AAV-CAG-Flex-GCaMP6s. In these mice, Cre is exclusively expressed in pyramidal neurons. For double labelling of GCaMP6 and a structural marker (tdTomato), two viruses (AAV2/1-CAG-Flex-GCaMP6s and AAV2/1-CAG-Flex-tdTomato; serotypes 2/1; $>2 \times 10^{13}$ genome copies per ml) were mixed at equal volumes and injected into L5 of the motor cortex. In experiments where L5 soma calcium activity was imaged, *Thy1*-Cre transgenic mice were used to label L5 pyramidal neurons specifically[44].

Motor cortex containing GCaMP-positive apical dendritic branches (from L5 neurons) at the depth of 0–200 μm below the pial surface were imaged for detecting the activities of tuft dendrites. In mice infected with viruses, to confirm that layer 2/3 were not labelled by GCaMP6, a glass window was placed over the region

of interest following the completion of imaging experiments, and imaging of layer 2/3 was performed to ensure the absence of GCaMP6 expression in layer 2/3 neurons.

In vivo two-photon imaging was performed with an Olympus Fluoview 1000 two-photon system (tuned to 920 nm) equipped with a Ti:Sapphire laser (MaiTai DeepSee, Spectra Physics). The average laser power on the tissue sample was ~20–30 and 50–60 mW for imaging in the L1 and L5 of the cortex respectively. All experiments were performed using a ×25 objective (numerical aperture 1.1) immersed in an ACSF solution and with a ×1.5 (soma) and ×4 (dendrites) digital zoom. All images were acquired at frame rates of 2 Hz (2-μs pixel dwell time) except for Fig. 1e and Extended Data Fig. 2d–f in which images were acquired at 4 Hz. Typical imaging window for Ca^{2+} imaging of tuft dendrites was 160 μm by 80 μm. Image acquisition was performed using FV10-ASW v.2.0 software and analysed post hoc using NIH ImageJ software.

Two-photon laser cutting of tuft dendrites of L5 pyramidal neurons. In Extended Data Fig. 4, dendrites expressing GCaMPs were cut by parking the two-photon laser beam on a small ROI that spanned the diameter of the dendrite for 5 s (tuned to ~890 nm and power was increased gradually until a sudden increase in fluorescence intensity was observed). After laser cut, a physical break could be observed between the two segments of dendrite. In the control experiment, the laser beam was parked ~30 μm away from the branch of interest without damaging it (Extended Data Fig. 4e).

Drugs applications. Approximately 200 μl of MK801 (M107, Sigma Aldrich), bicuculline (C2503, Tocris Bioscience), CaMKII inhibitors KN-93 (K1385, Sigma Aldrich)/KN-62 (I2142, Sigma Aldrich) or TTX citrate (1069; Tocris) was applied with a pipette to the surface of the primary motor cortex after removing a small bone flap (~200 μm in diameter) adjacent to a thinned skull window. CaMKII inhibitors KN-62 (100 μM) and KN-93 (100 μM) were first dissolved in dimethylsulphoxide (DMSO) and then in ACSF. The final DMSO concentration is lower than 1% (v/v) in ACSF. MK801 (100 μM), bicuculline (100 μM) and TTX (100 μM) were dissolved directly in ACSF to final concentrations. The bone flap for drug delivery was made during awake head mounting and covered with a silicone elastomer such that it could be easily removed at the time of imaging. Because small molecules diffuse rapidly in the cortex, we estimated that the drug concentration was reduced ~10 times in the imaged cortical region, such that the final effective concentration for CaMKII inhibitors, MK801, bicuculline and DMSO would be 10 μM, 10 μM and <0.1%, respectively. As the control experiment, we applied ACSF after removing a bone flap.

In experiments in which TTX and ACSF were injected to L1 and 5, we used a similar approach to that of virus injections (see above). In brief, TTX or ACSF was slowly injected to L1 or 5 (1 μl of 10 μM TTX in ACSF) with a glass microelectrode via pressure injections (Picospritzer III; 15 p.s.i., 12 ms, 0.8 Hz) 0.5 mm lateral to the forelimb motor region at a depth of 500–700 μm below the pial surface (Fig. 5a and Extended Data Fig. 9a, b). Mice were allowed to awake from weak isoflourane anaesthesia over the next 20 min. Ca^{2+} imaging was performed 20 min after TTX injection in L1 or L5 in the forelimb region of the motor cortex. To map the extent of local TTX diffusion, Texas Red dye was mixed and co-injected to the cortex. After imaging, mice were perfused with 4% paraformaldehyde (PFA) and their brains were postfixed overnight. Brains were sectioned with a vibratome at 100-μm sections. Confocal images were acquired using a Zeiss LSM 700 confocal microscope (×10 air objective, numerical aperture 0.3). The extent of Texas Red dye spread was estimated by the line at which the Texas Red fluorescence was less than 20% of its peak level. On the basis of this definition, we determined the fluorescence signals Texas Red spread to a region 237 ± 19 μm in L5 and 188 ± 10 μm in L1, respectively (Extended Data Fig. 9a, b).

Muscimol (Sigma) inactivation was achieved by thinning the skull overlying forelimb motor cortex and creating a small bone flap (~200 μm in diameter) where ~100 nl of muscimol (5 μg μl^{-1} in ACSF) or ACSF (control injection in forelimb region) was injected bilaterally (Picospritzer III; 15 p.s.i., 12 ms, 0.8 Hz) over 10–15 min into the motor cortex at depth of ~500–700 μm from pial surface, under isoflurane anaesthesia. The bone flap was then sealed and mice were allowed to recover in their home cage for >1 h before treadmill training. For control injections in a region other than forelimb motor cortex, muscimol was injected in the barrel cortex, using the coordinate of 1.3 mm posterior and 3.5 mm lateral from bregma.

SST neuron deletion and silencing. Deletion of SST interneurons was accomplished in *SST-IRES-Cre/Rosa26-stop-iDTR* mice by intraperitoneal injection of 1 μg diphtheria toxin (Sigma Aldrich) (three injections 12 h apart; diphtheria toxin diluted in saline; 1 μg diphtheria toxin/100 μl saline). *In vivo* two-photon imaging was performed to examine the loss of layer 2/3 SST interneurons in *SST-IRES-Cre/Rosa26-stop-iDTR* mice infected with AAV-CAG-Flex-GFP viruses before and 2 days after diphtheria toxin treatment. Deletion efficiency over two days was calculated by counting the number of SST GFP-positive somata in the same region of the forelimb motor cortex (from the pial surface to layer 2/3 (depth of approximately 250 μm)) in a 0.03 mm^2 cortical window) before and after diphtheria toxin treatment.

To examine the specificity of SST deletion and its effect on other cortical cells in motor cortex, we examined parvalbumin interneurons and microglia (Iba1) cell density by immunohistochemistry following 2 days after SST deletion. After imaging experiments, mice were anaesthetized and perfused with 20 ml PBS. The brain was removed and fixed for 1 h in 4% PFA at 4 °C. Tissue was rinsed three times with PBS, embedded in 2% agarose, and sectioned at 100 μm with a vibratome. Sections were permeabilized in 1% Triton X-100 in PBS for 3 h and blocked with 5% normal goat serum for 1 h. Sections were incubated overnight with primary antibodies against parvalbumin (Sigma, 1:300) and Iba1 (Wako, 1:500). Sections were then washed three times with PBS/0.05% Tween-20, and then incubated with Alexa Fluor-conjugated goat anti-rabbit IgG secondary antibodies (Life Technologies) 1:500 in PBS for 2 h. Sections were washed and mounted for imaging. Confocal images were obtained on a Zeiss LSM 700 confocal microscope. We found that there was no difference in the density (imaging area: parvalbumin, 1,250 × 1,250 μm^2; Iba1, 250 × 250 μm^2) of parvalbumin-positive cells or Iba1-positive microglia cells in forelimb motor cortex of mice deleted of SST cells as compared to control mice with SST cells ($P = 0.22$ and $P = 0.55$ for parvalbumin and Iba1 respectively; Extended Data Fig. 8c, d).

For acute SST interneuron silencing, Cre-dependent AAV2-hSyn-DIO-hM$_4$D(Gi)-mCherry was used to drive DREADD expression in *SST-IRES-Cre/Thy1*-GCaMP2.2c mice (AAV, serotype 2/1; ~10^{12} genome copies per ml; produced by UNC Vector Core). For validation of DREADD-hM$_4$Di silencing of SST interneuron activity, two viruses (AAV2/1-CAG-Flex-GCaMP6s and AAV2-hSyn-DIO-hM$_4$Di-mCherry) were mixed at equal volumes and injected into primary motor cortex of *SST-IRES-Cre* mice. Viruses were injected into the primary motor cortex two weeks before Ca^{2+} imaging. CNO (C0832, Sigma Aldrich) was dissolved in saline to a concentration of 0.5 mg ml^{-1}. Ca^{2+} imaging of dendritic tufts was performed during the initial running trials (forward and backward) without CNO. Twenty minutes after CNO was administrated by intraperitoneal injection to each mouse (0.3 ml per 30 g body weight), the same cortical region was re-imaged during a second session of running trials (Extended Data Fig. 8f).

It should be noted that the efficiency of CNO in activating hM$_4$Di was tested by expressing Flex-GCaMP6 and DIO-hM$_4$Di-mCherry in the same cell. We found that in SST neurons that were co-transfected with hM$_4$Di-mCherry and GCaMP6 viruses, the expression level of hM$_4$Di-mCherry was significantly lower than that in SST neurons transfected with hM$_4$Di-mCherry viruses only (Extended Data Fig. 8e). In the experiments in which dendritic Ca^{2+} spikes and the overlap index in pyramidal neurons were measured (Fig. 4c, d), SST neurons were transfected only with hM$_4$Di-mCherry viruses. We expect that the reduction of SST neuronal activity in those experiments would be higher (probably ~2-fold) than that shown in Fig. 4b.

Imaging data analysis. ROIs corresponding to visually identifiable spines, apical tuft dendrites or somata were selected for quantification. During running trials, the lateral movement of the images was typically less than 1 μm. Vertical movements were infrequent and minimized owing to flexible belt design, two micro-metal bars attached to the animal?s skull (described above) by dental acrylic, and a custom-built body support to minimize spinal cord movements generated by the hindlimbs. All imaging stacks were registered using ImageJ plugin StackReg. Active spines, dendrites and L5 pyramidal neurons that could be identified in all imaged sessions were included in the data set. The fluorescence time course of each spine, dendritic segment or cell body was measured with ImageJ software by averaging all pixels within the ROIs covering the dendrite or somata. The $\Delta F/F_0$ value was calculated as $\Delta F/F_0 = (F - F_0)/F_0 \times 100\%$, in which F_0 is the baseline fluorescence signal averaged over a 2-s period before the onset of the motor task. In all imaging sessions, resting awake images were collected before running trials.

Dendritic Ca^{2+} spikes were defined as the events when changes of fluorescence ($\Delta F/F_0$) observed in both dendritic spines and shaft (or dendritic branch) were >35% for GCaMP6s and >20% for GCaMP2.2c during the 2.5-min imaging sessions. F_0 is the fluorescence intensity in dendritic segments after background subtraction. Fluorescence background was measured from a region adjacent to the dendrite segment. The threshold for detecting dendritic spikes was more than three times the standard deviation of baseline fluorescence noise for GCaMP6s or GCaMP2.2c, respectively. As shown in Fig. 1h, most Ca^{2+} spikes were found to have a fluorescence increase of 100 to >1,000% for GCaMP6s and 50 to >200% for GCaMP2.2c.

In Extended Data Fig. 2f, Ca^{2+} spike rise time was determined by measuring the time between the threshold of detecting dendritic spikes to the peak amplitude. Decay time was measured as the time between the peak amplitude and the decay back to the threshold. The Ca^{2+} spike frequency was determined by recording the number of Ca^{2+} spikes over 2.5 min from the same motor region under various conditions (Extended Data Fig. 2h).

In Fig. 1i, task-specific Ca^{2+} spikes were defined as dendrites that showed activity (one or more events) in one task but not in the other running task (a 2.5-min imaging

session for each direction). In Fig. 4d, the overlap index was calculated by dividing the number of Ca^{2+} spikes on individual dendrites that activated during both forward/backward running trials by the number of Ca^{2+} spikes detected on individual dendrites that activated during forward or backward or both running trials.

Dendritic spine Ca^{2+} transients are defined as those with changes of fluorescence ($\Delta F/F_0$) >35% for GCaMP6s and >15% for GCaMP2.2c during the 2.5-min imaging sessions. The threshold is more than three times the standard deviation of baseline fluorescence noise for GCaMP6s ($\Delta F/F_0$: 33.9 ± 3.1%) and GCaMP 2.2c ($\Delta F/F_0$: 13.6 ± 3.7%). Spine Ca^{2+} transients detected with GCaMP6s were found to have an average fluorescence increase of 100.2 ± 3.8% ($\Delta F/F_0$) with 52% of spines exhibiting 1–4 transients per 2.5 min and 48% of spines exhibiting >5 transients per 2.5 min. In GCaMP2.2c-expressing dendrites, the average peak $\Delta F/F_0$ of Ca^{2+} transients in spines is 33 ± 0.9% during forward running trials. On average, active spines detected with 2.2c exhibited 3 transients per 2.5 min ($n = 65$ spines from 12 dendritic branches).

Ca^{2+} spikes and spine Ca^{2+} transients under analysis were always on the same branches. Consistent with previous studies[22], the ratio of Ca^{2+} fluorescence intensity in spine heads to neighbouring shafts for inactive spines was 0.46 ± 0.02 for GCaMP6s and 0.41 ± 0.02 for GCaMP2.2c (Extended Data Fig. 5k, l). For spines active at the time of spike generation and analysed in the present study, this ratio was 1.56 ± 0.13 for GCaMP6s and 1.71 ± 0.10 for GCaMP2.2c, significantly higher than that for neighbouring inactive spines (Extended Data Fig. 5k, l; $P < 0.001$). In Fig. 2, Ca^{2+} spike-induced potentiation of spines was determined by measuring the percentage change of the average peak $\Delta F/F_0$ of transients before and after Ca^{2+} spike generation. Most spines undergoing Ca^{2+} spike-induced potentiation had several spine transients before and after the spike. In Fig. 2, the peak $\Delta F/F_0$ of 3–4 Ca^{2+} transients from each spine was averaged before and after the spike to measure Ca^{2+} spike-induced changes. For active spines that did not experience a Ca^{2+} spike, 3–4 transients per spine were analysed for peak $\Delta F/F_0$ and compared between the first two and last two running trials. In Extended Data Fig. 5n, the peak $\Delta F/F_0$ of a single spine Ca^{2+} transient was compared immediately before and after the spike to measure Ca^{2+} spike-induced changes.

Spine diameter analyses was performed according to previous studies[47]. To correct for varying imaging conditions, the ratio of spine head diameter to adjacent dendritic shaft diameter was used to measure spine head diameter in tuft dendrites expressing both GCaMP6s and tdTomato. After background subtraction, the fluorescence intensity of the spine (the intensity of all pixels covering the spine in the best focal plane) was divided by the fluorescence intensity of the adjacent dendritic shaft. We measured spine head diameter before Ca^{2+} spike generation, 2 min and 40 min after spikes in active spines and non-active spines on the same dendritic segment (Fig. 2g, h). Spine size change was calculated by comparing spine size measurement 2 or 40 min after spikes to the first spine size measurement (before spikes).

Ca^{2+} activities in L5 trunk and soma were measured according to recently-published studies[20]. Task-specific apical trunk (Fig. 1j, l) and somata (Extended Data Fig. 3b–d) were defined as ROIs that showed activity (one or more transients) in one task but not in the other. In Fig. 5 and Extended Data Fig. 9c, e, GCaMP6s fluorescence changes over 20 and 40 min from individual L5 neurons was quantified as averaged $\Delta F/F_0$ over 2.5-min period of running after each training block.

Statistical analyses. All imaging and behavioural data were presented as mean ± s.e.m. Student?s t-test (two-tailed) was used to test for differences between groups whose distributions passed tests for normality (Kolmogorov–Smirnov). Wilcoxon matched-pairs signed rank test and Mann–Whitney U test were used to analyse those groups whose distributions did not pass tests for normality. Significant levels were set at $P \leq 0.05$. All statistical analyses were performed using the GraphPad Prism. Statistical details for the experiments described in Figs 1–5 are detailed below, including the sample size (number of dendrites, spines, neurons and number of mice). Sample sizes were chosen to ensure adequate power with the statistical tests while minimizing the number of animals used in compliance with ethical guidelines. Experiments did not involve randomization or blinding because no dendritic spikes, potentiated or depotentiated spines, or behavioural performance of animal groups were predefined.

For comparing changes in steady run gait during training in Fig. 1c, the distance between adjacent footprints was measured with ink over 5 trials before and after two 2-min training blocks from 9 mice. Normalized average stride length: 1 ($t = 0$–5 min), 1.3 ± 0.06 ($t = 25$–30 min, $P = 0.0003$, paired Student?s t-test), 1.6 ± 0.07 ($t = 50$–55 min, $P = 0.0004$, paired Student?s t-test).

The number of apical tuft Ca^{2+} spikes during forward or backward running in Fig. 1f was compared to quiet resting state in L3 mice expressing GCaMP2.2c. Fold change was calculated from the average number of spikes over 2.5 min recording within an imaging window of 160 μm by 80 μm. Forward run: 24.5 ± 2.4 versus quiet resting: 2.4 ± 0.4, $P = 2.5 \times 10^{-7}$, paired Student?s t-test. Backward run: 19.7 ± 1.7 versus quiet resting, $P = 2.6 \times 10^{-7}$, paired Student?s t-test.

Distribution of Ca^{2+} spike peak amplitude in Fig. 1h was measured during forward running. GCaMP6s: $n = 141$ spikes from 9 mice; peak amplitude was 945 ± 51%. GCaMP2.2c: $n = 213$ spikes from 5 mice; peak amplitude was 111.9 ± 3.9%. GCaMP2.2c with local MK801 applied to layer 1: $n = 31$ spikes from 4 mice; peak amplitude was 42.7 ± 3.4%. The peak amplitude of Ca^{2+} spikes detected by GCaMP2.2c is significantly reduced in the presence of MK801 ($P < 0.0001$; Mann–Whitney test).

Mean percentage change in the peak amplitude of dendritic spine Ca^{2+} transients in Fig. 2c was measured under various conditions. Forward running spike-induced potentiation: 74.7 ± 7.3% increase in peak $\Delta F/F_0$ ($n = 80$ spines from 9 mice (G6s) and 12 mice (G2.2c), $P < 0.001$, Wilcoxon matched-pairs signed rank test; also see Extended Data Fig. 5m, n). Backward running spike-induced potentiation: 52.7 ± 10.9% increase in peak $\Delta F/F_0$ ($n = 18$ spines from 4 mice, $P = 0.0002$, paired Student?s t-test). Spines not active at the spike generation: 1.8 ± 5.8% increase in peak $\Delta F/F_0$ ($n = 22$ spines from 5 mice, $P = 0.76$, paired Student?s t-test). Active spines with no encounter of spikes: 4.2 ± 2.9% increase in peak $\Delta F/F_0$ ($n = 35$ spines from 9 mice, $P = 0.15$, paired Student?s t-test). Local application of MK801: 0.2 ± 3.5% increase in peak $\Delta F/F_0$ ($n = 31$ spines from 4 mice, $P = 0.95$, paired Student?s t-test). Local KN-93: 6.2 ± 5.3% increase in peak $\Delta F/F_0$ ($n = 20$ spines from 3 mice, $P = 0.26$, paired Student?s t-test). Local KN-62: 3.9 ± 5.4% increase in peak $\Delta F/F_0$ ($n = 18$ spines from 3 mice, $P = 0.48$, paired Student?s t-test). ACSF control: 39.5 ± 11.3% increase in peak $\Delta F/F_0$ ($n = 17$ spines from 2 mice, $P = 0.003$, paired Student?s t-test). The degrees of spine potentiation under various conditions were compared and shown in Fig. 2c (unpaired Student?s t-test).

Correlation of peak amplitude of spine Ca^{2+} transients and peak amplitude of Ca^{2+} spike for dendrites expressing GCaMP6s ($n = 30$, $R = 0.61$, Pearson correlation coefficient, $P = 0.0003$) and GCaMP2.2c ($n = 51$, $R = 0.68$, Pearson correlation coefficient, $P < 0.0001$) in Fig. 2d.

Correlation of peak amplitude of spine Ca^{2+} transients and the spine head and shaft fluorescence ratio for dendrites expressing GCaMP6s ($n = 30$, $R = 0.65$, Pearson correlation coefficient, $P = 0.0001$) and GCaMP2.2c ($n = 51$, $R = 0.73$, Pearson correlation coefficient, $P < 0.0001$) in Fig. 2e.

The mean percentage change in the peak amplitude of spine Ca^{2+} transients in Fig. 2f was measured over 40-min training in spines active at the time of spike generation and in spines not active at the time of spike generation. Ca^{2+} spikes induced potentiation of spine Ca^{2+} transients: 53.1 ± 10.3% (0–2 min, $P = 2.2 \times 10^{-5}$, paired Student?s t-test), 51.8 ± 8.9% (20 min, $P = 3.6 \times 10^{-6}$, paired Student?s t-test), 47.6 ± 10.5% (40 min, $P = 0.0003$, paired Student?s t-test) ($n = 28$ spines from 6 mice). Non-active spines at spike: −2.0 ± 5.4% (0–2 min, $P = 0.67$, paired Student?s t-test), 8.6 ± 5.3% (20 min, $P = 0.08$, paired Student?s t-test), −2.6 ± 4.7% (40 min) ($n = 17$ spines, $P = 0.53$, paired Student?s t-test).

In tuft dendrites co-expressing GCaMP6s and tdTomato in Fig. 2h, spines undergoing Ca^{2+} spike-induced Ca^{2+} potentiation were measured for spine size changes at 40 min: 32.3 ± 11.1% ($n = 8$ spines from 5 mice, $P = 0.02$, paired Student?s t-test); at 2 min: 2.1 ± 7.1% ($n = 11$ spines from 11 mice, $P = 0.77$, paired Student?s t-test). Spine size change of inactive spines was measured at 40 min: 4.0 ± 7.6% ($n = 14$ spines from 5 mice, $P = 0.61$, paired Student?s t-test); at 2 min: 0.6 ± 4.5% ($n = 23$ spines from 11 mice, $P = 0.89$, paired Student?s t-test). The size change between active and inactive spines was also compared (unpaired Student?s t-test) and shown in Fig. 2h.

The mean percentage change in peak amplitude of spine Ca^{2+} transients in Fig. 3c was measured in forward running periods in this task-switching task (forward–backward–forward). Spines active <5 s before the spike induced during backward running: −38.6 ± 4.1% change in spine peak $\Delta F/F_0$ (ratio of the fluorescent intensity between spine heads and neighbouring shafts was 0.53 ± 0.03 during backward running-induced spike, $n = 18$ spines from 8 mice, $P = 4.1 \times 10^{-5}$, paired Student?s t-test); Spines active >5 s before the spike: −4.9 ± 3.4% decrease in peak $\Delta F/F_0$ ($n = 31$ spines from 8 mice, $P = 0.16$, paired Student?s t-test); Active spines without encountering spikes: 1.8 ± 5.0% increase in peak $\Delta F/F_0$ ($n = 31$ spines from 5 mice, $P = 0.66$, Wilcoxon matched-pairs signed rank test); Local application of MK801 in layer 1: −3.7 ± 3.2% decrease in peak $\Delta F/F_0$ ($n = 34$ from 3 mice, $P = 0.26$, Wilcoxon matched-pairs signed rank test). The percentage change between active and inactive spines was also compared (unpaired Student?s t-test) and shown in Fig. 3c.

No correlation between the percentage change in the peak amplitude of spine Ca^{2+} transients and dendritic Ca^{2+} spike amplitude in Fig. 3d ($n = 31$ spines, $R = -0.17$, $P = 0.35$).

In vivo two-photon imaging of the same cortical region in Fig. 4a reveal a substantial loss of layer 2/3 SST interneurons 2 days after diphtheria toxin treatment in *SST-IRES-Cre/Rosa26-stop-iDTR* mice infected with AAV-CAG-Flex-GFP viruses. Total cell number before deletion: 877 cells; total cell number after deletion: 222 cells ($n = 5$ mice, $P = 2 \times 10^{-5}$, paired Student?s t-test).

In vivo analysis in Fig. 4b of DREADD-induced silencing of SST neurons 20 min after CNO treatment. To assess the change in Ca^{2+} response after CNO treatment in layer 2/3 SST interneurons, averaged calcium activity over 1 min during quiet resting state was measured before and after CNO application (before: 1; after: 0.8 ± 0.05; $n = 27$ somata from 4 mice, $P < 0.0001$, Wilcoxon matched-pairs signed rank test).

Control mice and mice with SST neuron deletion (SST/DTR) or silencing (SST/hM_4Di) were subjected to both forward and backward running and Ca^{2+} spikes were recorded in layer 1 over 5 trials in each direction in Fig. 4d. Overlap index was calculated by dividing the number of Ca^{2+} spikes on branches that activated during both forward/backward running trials by the number of Ca^{2+} spikes detected on branches that activated during forward or backward or both running trials. SST-intact overlap index: $9.8 \pm 1.2\%$ ($n = 12$ mice, 107 overlapping spikes out of 987 spikes); SST-intact + diphtheria toxin overlap index: $7.1 \pm 2.3\%$ ($n = 3$ mice, 13 overlapping spikes out of 197 spikes); SST/DTR + diphtheria toxin overlap index: $42.3 \pm 4.0\%$ ($n = 8$ mice, 642 overlapping spikes out of 1,470 spikes), significantly higher than SST-intact + diphtheria toxin overlap index ($P = 0.0055$, Mann–Whitney *U* test); SST/hM_4Di + saline overlap index: 4.5 ± 2.4 ($n = 6$ mice, 14 overlapping spikes out of 215 spikes) SST/hM_4Di + CNO overlap index: 39.5 ± 6.3 ($n = 6$ mice, 331 overlapping spikes out of 757 spikes, $P = 0.002$, paired Student?s *t*-test).

In Fig. 4i, the mean percentage change of Ca^{2+} transient peak amplitude in previously forward-running potentiated spines after either forward or backward running in control and SST-deleted mice. Percentage change in spine $\Delta F/F_0$ in SST-intact control mice, after forward running: $4.2 \pm 2.9\%$ ($n = 35$ spines, $P = 0.15$, paired Student?s *t*-test), and after backward running: $5.7 \pm 5.6\%$ ($n = 18$ spines, $P = 0.32$, paired Student?s *t*-test). Percentage change in spine $\Delta F/F_0$ in SST-deleted mice, after forward running (potentiated): $14.9 \pm 6.1\%$ ($n = 24$ spines, $P = 0.02$, paired Student?s *t*-test), and after backward running (depotentiated): $-20.5 \pm 10.0\%$ ($n = 34$ spines, $P = 0.016$, Wilcoxon matched-pairs signed rank test).

Mean number of tuft Ca^{2+} transients over a 30-s running trial was measured in transgenic mice expressing GCaMP2.2c in Fig. 5b. ACSF to L1: 11.6 ± 0.8 ($n = 46$ trials from 3 mice); TTX to L1: 0.38 ± 0.1 ($n = 40$ trials from 3 mice, $P < 0.0001$, Mann–Whitney *U* test); TTX to L5: 20.4 ± 0.8 ($n = 61$ trials from 3 mice, $P < 0.0001$, Mann–Whitney *U* test); mean number of trunk Ca^{2+} transients over a 30-s running trial was measured in mice expressing GCaMP2.2c. ACSF to L1: 16.8 ± 1.8 ($n = 28$ trials from 5 mice); TTX to L1: 1.8 ± 0.2 ($n = 32$ trials from 3 mice, $P < 0.0001$, Mann–Whitney *U* test); TTX to L5: 13.4 ± 1.1 ($n = 36$ trials from 5 mice), comparable to ACSF to L1: 16.8 ± 1.8 ($P = 0.2771$, Mann–Whitney *U* test).

Normalized peak $\Delta F/F_0$ of L5 somata over a 30-s running trial in Fig. 5c. ACSF applied to L1: $124.5 \pm 6.1\%$ $\Delta F/F_0$ ($n = 123$ somata from 5 mice); TTX to L1: $52.1 \pm 2.5\%$ $\Delta F/F_0$ ($n = 133$ from 3 mice, $P < 0.0001$, Mann–Whitney *U* test); TTX to L5: $14.6 \pm 4.8\%$ $\Delta F/F_0$ ($n = 60$ from 2 mice), significantly lower than ACSF to L1: $124.5 \pm 6.1\%$ ($P < 0.0001$, Mann–Whitney *U* test).

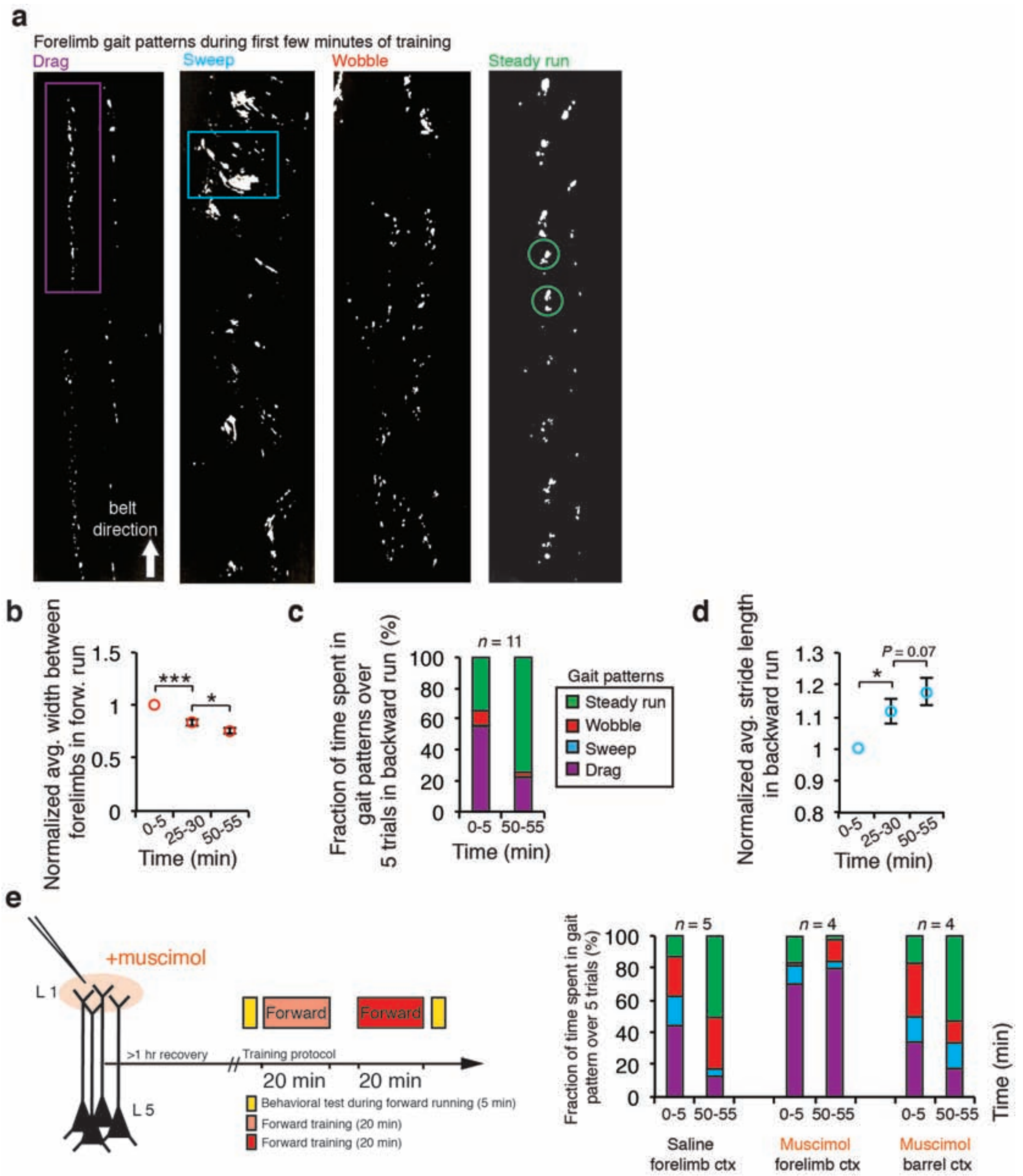
In Fig. 5e, the mean percentage change in peak amplitude of the apical trunk nexus in control and SST-deleted mice after the initial forward running (SST-intact: $63.9 \pm 8.4\%$, $n = 64$ trunk from 5 mice, $P = 3.1 \times 10^{-11}$, paired Student?s *t*-test; SST-deleted: $82.3 \pm 15.8\%$, $n = 64$ from 3 mice, $P < 0.0001$, Wilcoxon matched-pairs signed rank test) and the second forward running session (SST-intact: $21.9 \pm 10.6\%$, $P = 0.26$, Wilcoxon matched-pairs signed rank test; SST-deleted: $23.7 \pm 8.6\%$, $P = 0.009$, paired Student?s *t*-test). Backward running after the initial forward running session reduced Ca^{2+} activity at the nexus in SST-deleted mice (SST-intact: $1.0 \pm 9.2\%$, $n = 107$ from 3 mice, $P = 0.1181$, Wilcoxon matched-pairs signed rank test; SST-deleted: $-24.2 \pm 7.1\%$, $n = 77$ from 4 mice, $P = 0.001$, paired Student?s *t*-test). There is a significant reduction in the trunk Ca^{2+} activity after backward running in SST-deleted mice than in SST-intact mice ($P = 0.0063$, Mann–Whitney *U* test).

In Fig. 5f, the mean percentage change in the average $\Delta F/F_0$ of L5 somata after initial (SST-intact: $23.5 \pm 2.6\%$, $n = 113$ from 6 mice, $P < 0.0001$, Wilcoxon matched-pairs signed rank test; SST-deleted: $44.4 \pm 8.6\%$, $n = 74$ from 7 mice, $P = 0.0003$, Wilcoxon matched-pairs signed rank test) and second sessions (SST-intact: $15.3 \pm 3.4\%$, $P < 0.0001$, Wilcoxon matched-pairs signed rank test; SST-deleted: $17.9 \pm 3.6\%$, $P = 9.45 \times 10^{-6}$, paired Student?s *t*-test) of forward running in control and SST-deleted mice. Backward running after the initial forward running session reduced Ca^{2+} activity at L5 soma in SST-deleted mice, but not in control mice (SST-intact: $-1.0 \pm 1.6\%$, $n = 63$ from 3 mice, $P = 0.1038$, Wilcoxon matched-pairs signed rank test; SST-deleted: $-19.8 \pm 3.3\%$, $n = 67$ from 3 mice, $P = 0.0007$, paired Student?s *t*-test). There is a significant reduction in L5 soma activity after backward running in SST-deleted mice than in SST-intact mice ($P < 0.0001$, Mann–Whitney *U* test).

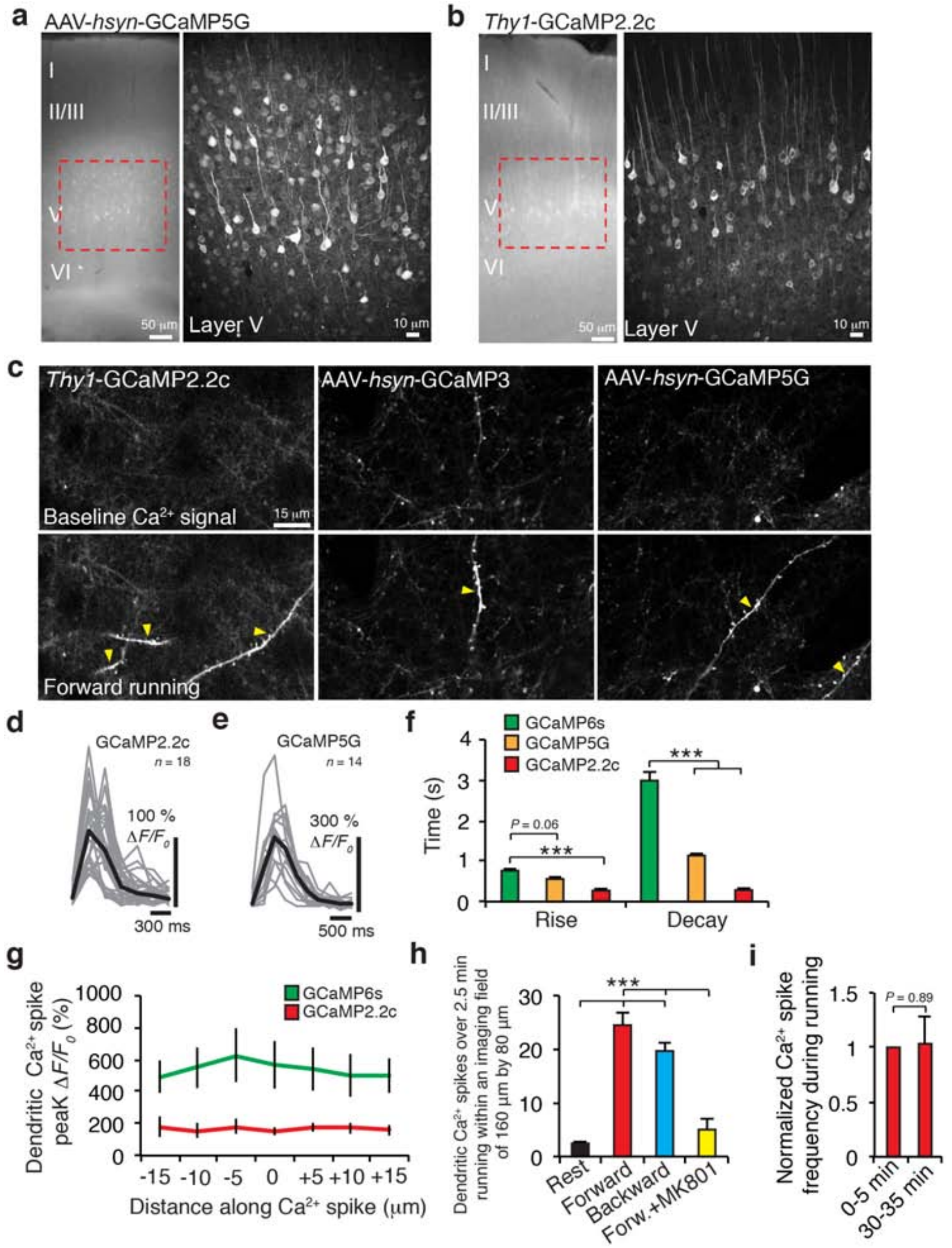
In Fig. 5g, the mean percentage change in stride length after initial (SST-intact: $35.7 \pm 4.0\%$, $P = 1.9 \times 10^{-7}$, paired Student?s *t*-test; SST-deleted: $34.4 \pm 4.9\%$, $P = 3.6 \times 10^{-6}$, paired Student?s *t*-test) and second sessions (SST-intact: $22.8 \pm 3.9\%$ $P = 0.0003$, paired Student?s *t*-test; SST-deleted: $17.9 \pm 3.8\%$ $P = 0.005$, paired Student?s *t*-test) of forward running in control and SST-deleted mice. Backward running after the initial forward running session reduced stride length in SST-deleted mice, not in SST-intact controls (SST-intact: $14.5 \pm 5.8\%$ $P = 0.14$, paired Student?s *t*-test; SST-deleted: $-25.9 \pm 6.03\%$, $P = 0.0004$, paired Student?s *t*-test). There is a significant reduction in stride length after backward running in SST-deleted mice than in SST-intact mice ($P = 0.0001$, unpaired Student?s *t*-test). The number of animals used under various conditions is shown in Fig. 5g.

In Fig. 5h, the mean percentage change in the width between forelimbs after initial (SST-intact: $-11.2 \pm 1.9\%$, $P = 1.3 \times 10^{-5}$, paired Student?s *t*-test; SST-deleted: $-17.6 \pm 2.1\%$, $P = 1.5 \times 10^{-6}$, paired Student?s *t*-test) and second (SST-intact: $-4.6 \pm 1.8\%$ $P = 0.038$, paired Student?s *t*-test; SST-deleted: $-4.2 \pm 3.3\%$, $P = 0.4375$, Wilcoxon matched-pairs signed rank test) sessions of forward running in control and SST-deleted mice. Backward running after the initial forward running session increased width distance in SST-deleted mice, not in SST-intact controls (SST-intact: $-4.7 \pm 1.9\%$, $P = 0.05$, paired Student?s *t*-test; SST-deleted: $7.5 \pm 3.5\%$, $P = 0.04$, paired Student?s *t*-test). There is a significant increase in width distance after backward running in SST-deleted mice when compared to that in SST-intact mice ($P = 0.004$, paired Student?s *t*-test).

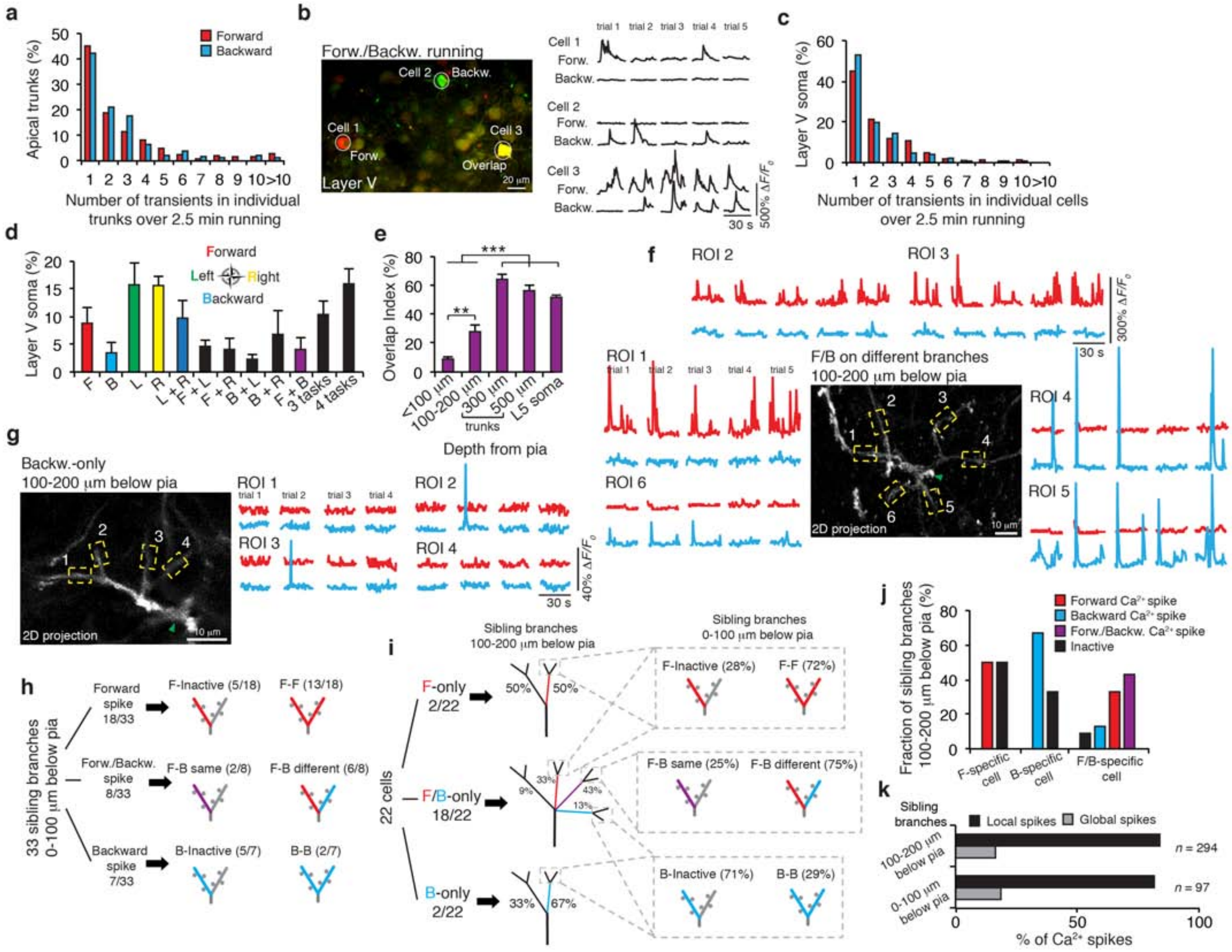
44. Chen, Q. *et al.* Imaging neural activity using Thy1-GCaMP transgenic mice. *Neuron* **76,** 297–308 (2012).
45. Yang, G., Pan, F., Chang, P. C., Gooden, F. & Gan, W. B. Transcranial two-photon imaging of synaptic structures in the cortex of awake head-restrained mice. *Methods Mol. Biol.* **1010,** 35–43 (2013).
46. Tennant, K. A. *et al.* The organization of the forelimb representation of the C57BL/6 mouse motor cortex as defined by intracortical microstimulation and cytoarchitecture. *Cereb. Cortex* **21,** 865–876 (2011).
47. Grutzendler, J., Kasthuri, N. & Gan, W. B. Long-term dendritic spine stability in the adult cortex. *Nature* **420,** 812–816 (2002).



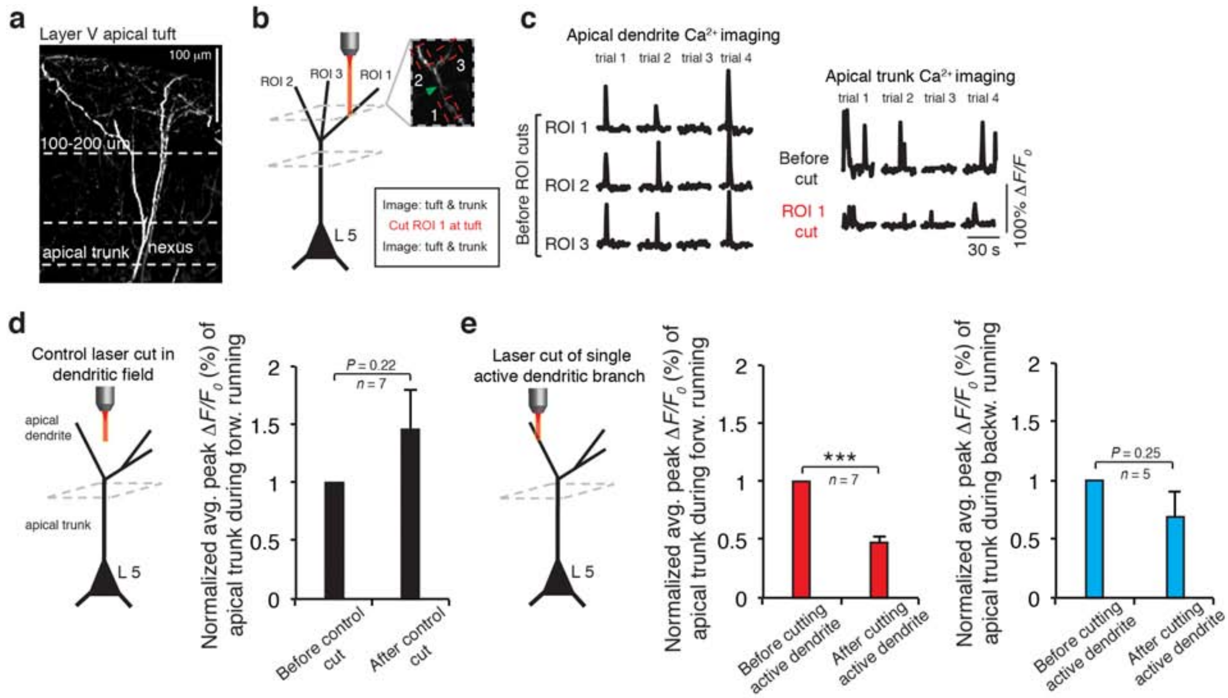
Extended Data Figure 1 | The forelimb motor cortex is important for treadmill running and performance improvement. **a**, Representative forelimb gait traces from wild-type mice running forward on the treadmill during the first few minutes of training. Several gait patterns (drag, sweep, wobble and steady run) were observed. **b**, Average width between forelimbs in steady run decreased during forward treadmill running. **c**, Backward running elicited changes in gait patterning over two 20-min training sessions. Initially, the mice exhibited mostly steady run (34%) and drag (55%) gait patterns without sweep. With continued training, mice refined their gait from drag to steady run (75%) over 40 min. **d**, Average stride length in steady run increased during backward treadmill training. **e**, Bilateral injections of muscimol, a GABA receptor agonist, into the forelimb motor cortex acutely impaired treadmill running performance ($n = 4$). Muscimol injected mice displayed high percentages of untrained gait features (drag: 70% (0–5 min) and 80% (50–55 min)), whereas mice injected with saline to forelimb motor cortex did not (drag: 44% (0–5 min) and 13% (50–55 min)) ($n = 5$). Muscimol injections into barrel cortex did not impair treadmill running performance (drag: 34% (0–5 min) and 18% (50–55 min)) ($n = 4$). Data are presented as mean ± s.e.m. $^{*}P < 0.05$, $^{***}P < 0.001$, paired t-test.



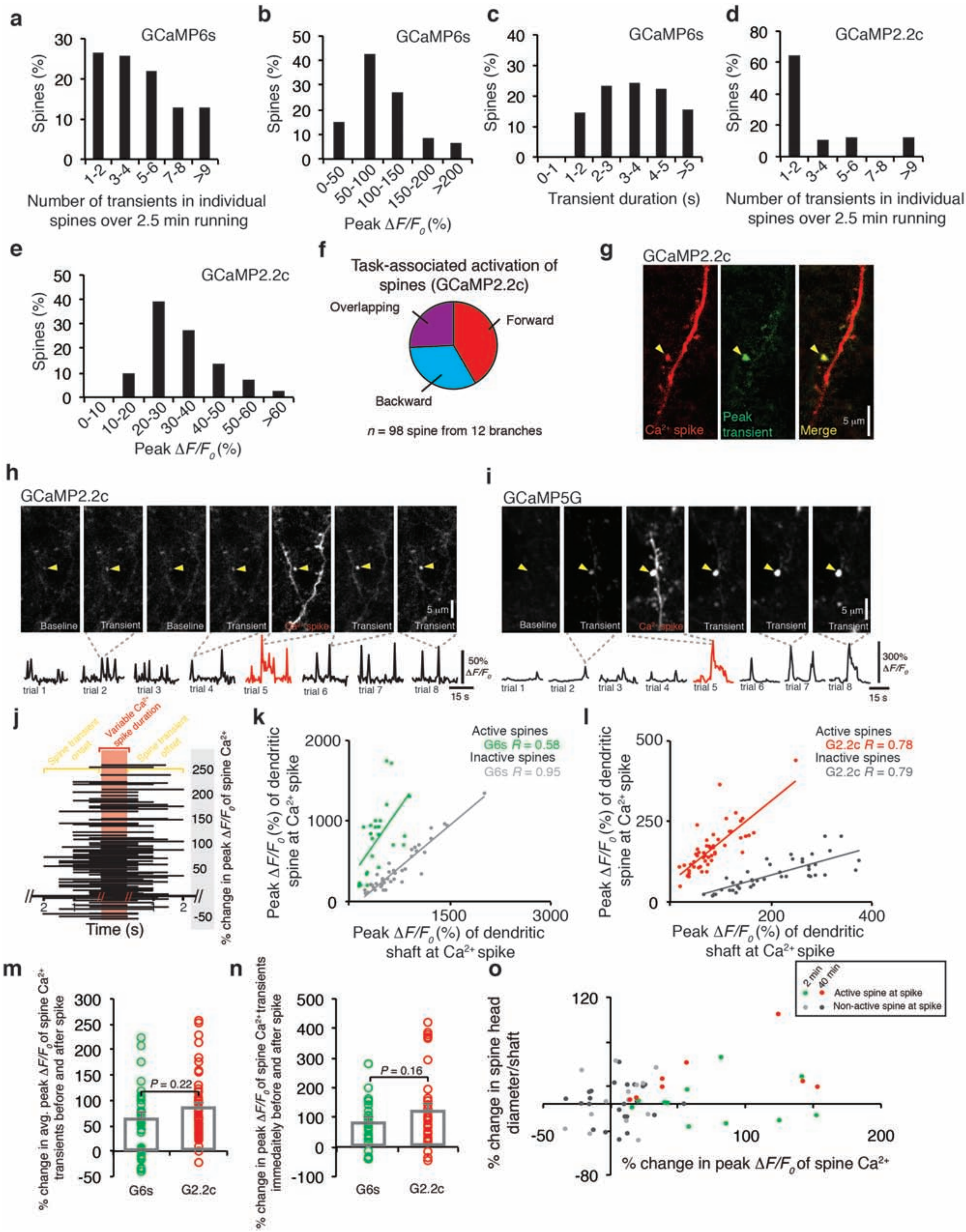
Extended Data Figure 2 | Detection of motor learning-induced Ca^{2+} spikes by various GCaMPs in apical tuft dendrites of layer 5 neurons in the motor cortex. **a**, **b**, Coronal sections of forelimb motor cortex from mice expressing AAV-GCaMP5G (**a**) or GCaMP2.2c (**b**). Boxed regions shows the expression of GCaMPs in L5. **c**, Two-photon images of GCaMP2.2c-, 3- and 5G-expressing dendrites during quiet resting state and forward running. Images of baseline Ca^{2+} signals under quiet resting state (top) and running-induced Ca^{2+} spikes (bottom; yellow arrowheads) are shown. **d**, **e**, Fast-scanning of apical tuft dendrites during forward running in mice expressing GCaMP2.2c ($n = 18$) and GCaMP5G ($n = 14$). Grey traces are individual Ca^{2+} transients and black trace represents the average. **f**, Average rise and decay times of apical tuft Ca^{2+} spikes during running for different GCaMPs (unpaired *t*-test). **g**, Measurements of Ca^{2+} fluorescence along long dendritic segments in the plane of imaging. Both GCaMP6s ($n = 10$) and GCaMP2.2c ($n = 7$) detected comparable fluorescent signals across entire dendritic segments. **h**, Number of Ca^{2+} spikes during quiet resting, running forward, running backward, and with local application of MK801(paired *t*-test). Ca^{2+} spikes were detected on L5 tuft branches in an image field (160 × 80 μm) over 2.5 min in *Thy-1* GCaMP2.2c transgenic mice. **i**, The number of dendritic Ca^{2+} spikes generated in early running trials was not significantly different from that in later (30 min) running trials ($P = 0.89$, paired *t*-test). Scale bar, 50 μm (**a**, **b**) and 15 μm (**c**). $^{*}P < 0.05$, $^{***}P < 0.001$.



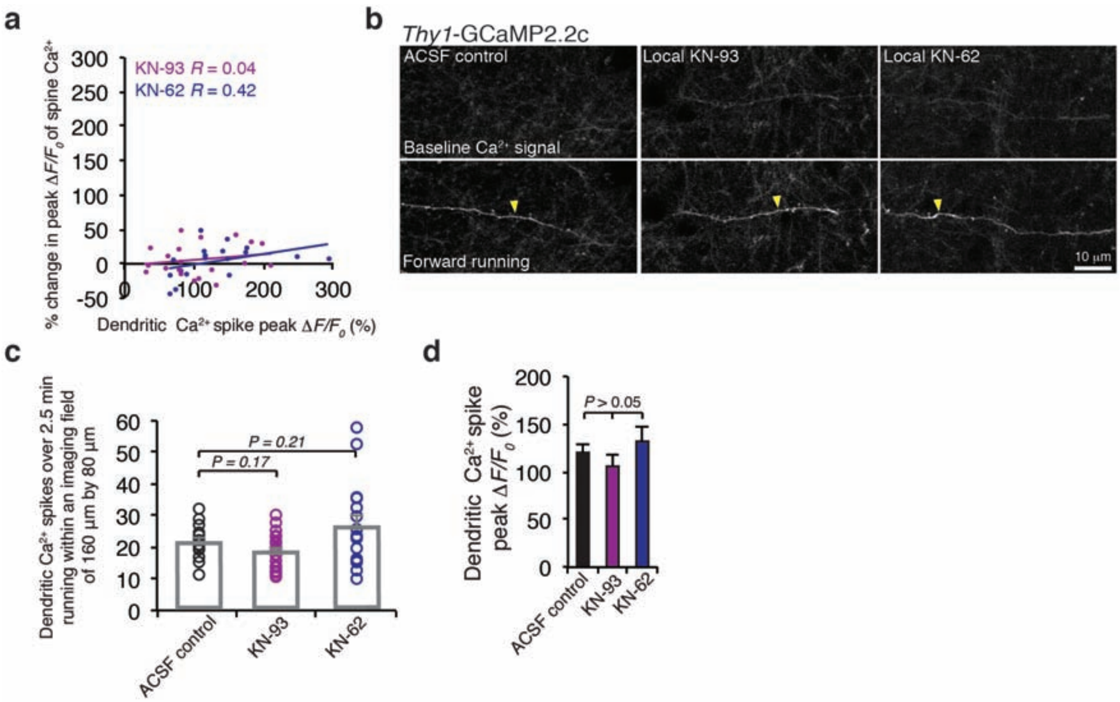
Extended Data Figure 3 | The overlap of Ca^{2+} activity in apical dendritic trunk, somata, and apical tuft branches of L5 pyramidal neurons in the motor cortex in response to treadmill running. **a**, Frequency distribution of the number of Ca^{2+} transients detected over 2.5 min in individual apical trunks in mice running forward and backward. **b**, Ca^{2+} imaging of layer 5 somata during forward and backward running. **c**, Frequency distribution of the number of Ca^{2+} transients in individual L5 somata detected over 2.5 min in mice running forward and backward ($n = 504$ cells from 10 mice). **d**, Layer 5 somata responded to multiple tasks when running in four directions ($n = 242$ cells from 4 mice). **e**, Summary of the overlap of Ca^{2+} activity in response to forward and backward running across different cortical layers in the motor cortex. **f**, Two-dimensional projection of six apical tuft branches from the same L5 neuron expressing GCaMP6s. Six ROIs corresponding to different branches were analysed over 2.5 min forward or backward running. Green arrowhead marks the location of the trunk (~200 μm below the pia). Note that there is little or no overlap between forward spikes and backward spikes in these branches. **g**, Two-dimensional two-photon image of four apical tuft branches expressing GCaMP2.2c from an individual L5 neuron. Green arrowhead marks the trunk. Four ROIs corresponding to different branches were analysed over four trials of forward and backward running. ROI 2 and 3 generated Ca^{2+} transients in response to backward running. **h**, Distribution of forward and backward running-induced Ca^{2+} spikes on 33 sibling branches located 0–100 μm below the pia. Data were analysed over five 30-s trials of forward and backward running. **i**, **j**, Distribution of forward and backward running-induced Ca^{2+} spikes on 80 sibling branches located 100–200 μm below the pia (22 cells). Approximately 43% branches at this cortical depth exhibited spikes in response to both forward and backward running. Each of these branches was connected to higher order branches that were either inactive or exhibited spikes in response to forward, backward or both running modes (dotted boxes in **i**). **k**, Percentage of local or global Ca^{2+} spikes observed on sibling branches located at two different depths below the pia. $^{**}P < 0.01$, $^{***}P < 0.001$, unpaired t-test.



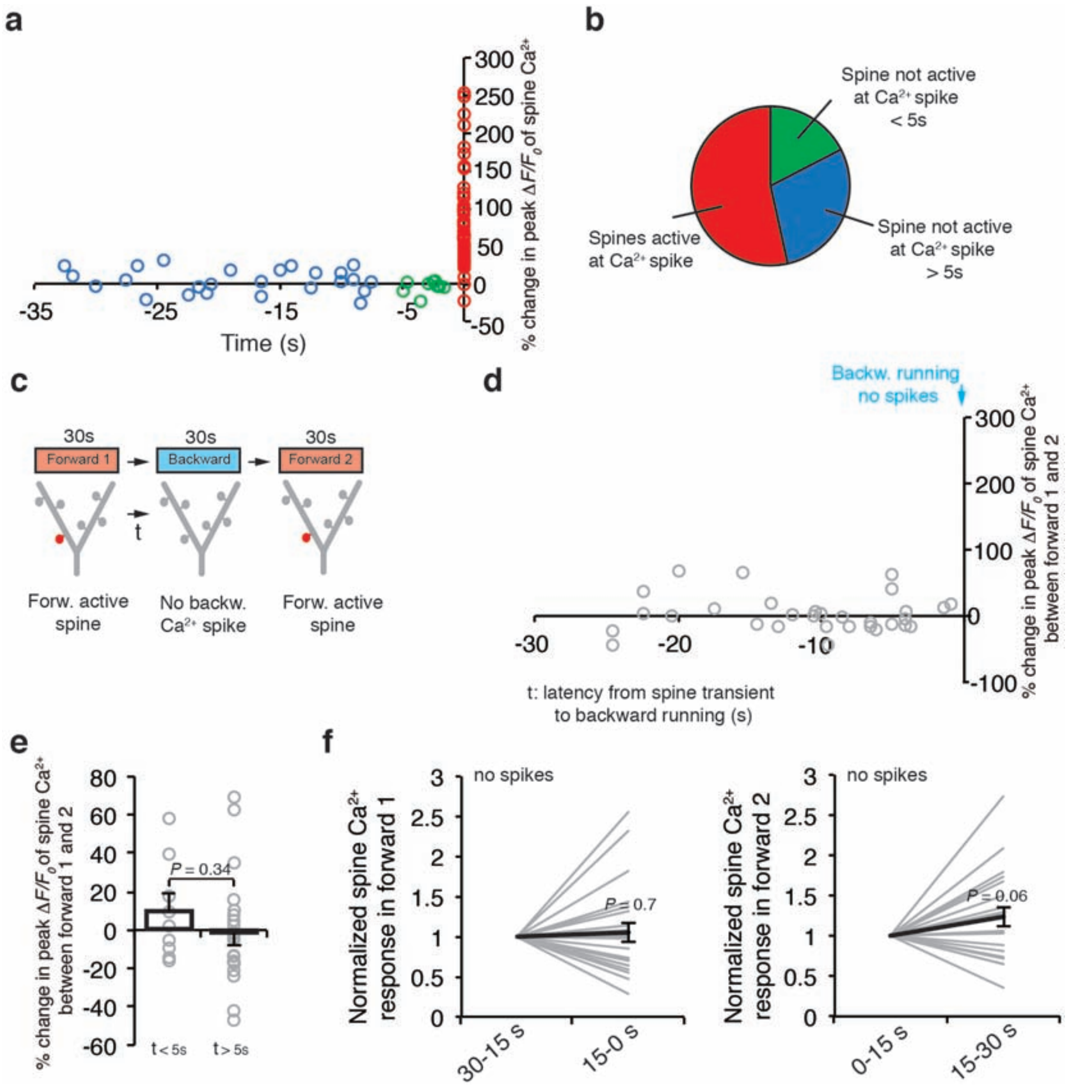
Extended Data Figure 4 | *In vivo* two-photon laser cutting of one apical tuft branch reduces calcium activity at the apical trunk. **a**–**c**, Example of an individual neuron with three apical tuft branches (at 150 μm) and the apical trunk (at 300 μm; **a**) showed Ca^{2+} spikes during forward running. Laser cutting one tuft branch (**b**) induced a significant reduction in average peak $\Delta F/F_0$ of the trunk during forward running (**c**). **d**, Summary of average peak $\Delta F/F_0$ of the trunk before and after parking the laser beam in a region 30 μm away from the active dendrite (control cut; $n = 7$, $P = 0.22$, paired t-test). There was no significant change in the activity of the trunk in this control experiment. **e**, Summary of average peak $\Delta F/F_0$ of the trunk before and after cutting an active dendrite during forward ($n = 7$, $P < 0.001$, paired t-test) and backward running ($P = 0.25$; $n = 5$). ***$P < 0.001$. As expected, cutting a dendritic branch exhibiting forward running-induced Ca^{2+} spikes reduced the activity of the trunk when animals ran in the forward direction (left). When the branch with forward Ca^{2+} spikes was cut, the average activity of the trunk was also reduced for the backward direction (right), even though the uncut branch still exhibited Ca^{2+} spikes in response to backward running. This is probably related to the fact that apical tufted branches possess spines that are active during both forward and backward running (data not shown). Cutting a branch eliminated the contribution of not only dendritic Ca^{2+} spikes but also synaptic inputs to the depolarization at the trunk.


a
GCaMP6s
Spines (%)
Number of transients in individual spines over 2.5 min running
b
GCaMP6s
Spines (%)
Peak ΔF/F0 (%)
c
GCaMP6s
Spines (%)
Transient duration (s)
d
GCaMP2.2c
Spines (%)
Number of transients in individual spines over 2.5 min running
e
GCaMP2.2c
Spines (%)
Peak ΔF/F0 (%)
f
Task-associated activation of spines (GCaMP2.2c)
Overlapping
Forward
Backward
n = 98 spine from 12 branches
g
GCaMP2.2c
Ca2+ spike
Peak transient
Merge
5 μm
h
GCaMP2.2c
Baseline
Transient
Ca2+ spike
5 μm
trial 1
trial 2
trial 3
trial 4
trial 5
trial 6
trial 7
trial 8
50% ΔF/F0
15 s
i
GCaMP5G
300% ΔF/F0
j
Spine transient onset
Variable Ca2+ spike duration
Spine transient offset
Time (s)
% change in peak ΔF/F0 of spine Ca2+
k
Active spines
G6s R = 0.58
Inactive spines
G6s R = 0.95
Peak ΔF/F0 (%) of dendritic spine at Ca2+ spike
Peak ΔF/F0 (%) of dendritic shaft at Ca2+ spike
l
Active spines
G2.2c R = 0.78
Inactive spines
G2.2c R = 0.79
m
% change in avg. peak ΔF/F0 of spine Ca2+ transients before and after spike
P = 0.22
G6s
G2.2c
n
% change in peak ΔF/F0 of spine Ca2+ transients immediately before and after spike
P = 0.16
o
2 min
40 min
Active spine at spike
Non-active spine at spike
% change in spine head diameter/shaft
% change in peak ΔF/F0 of spine Ca2+

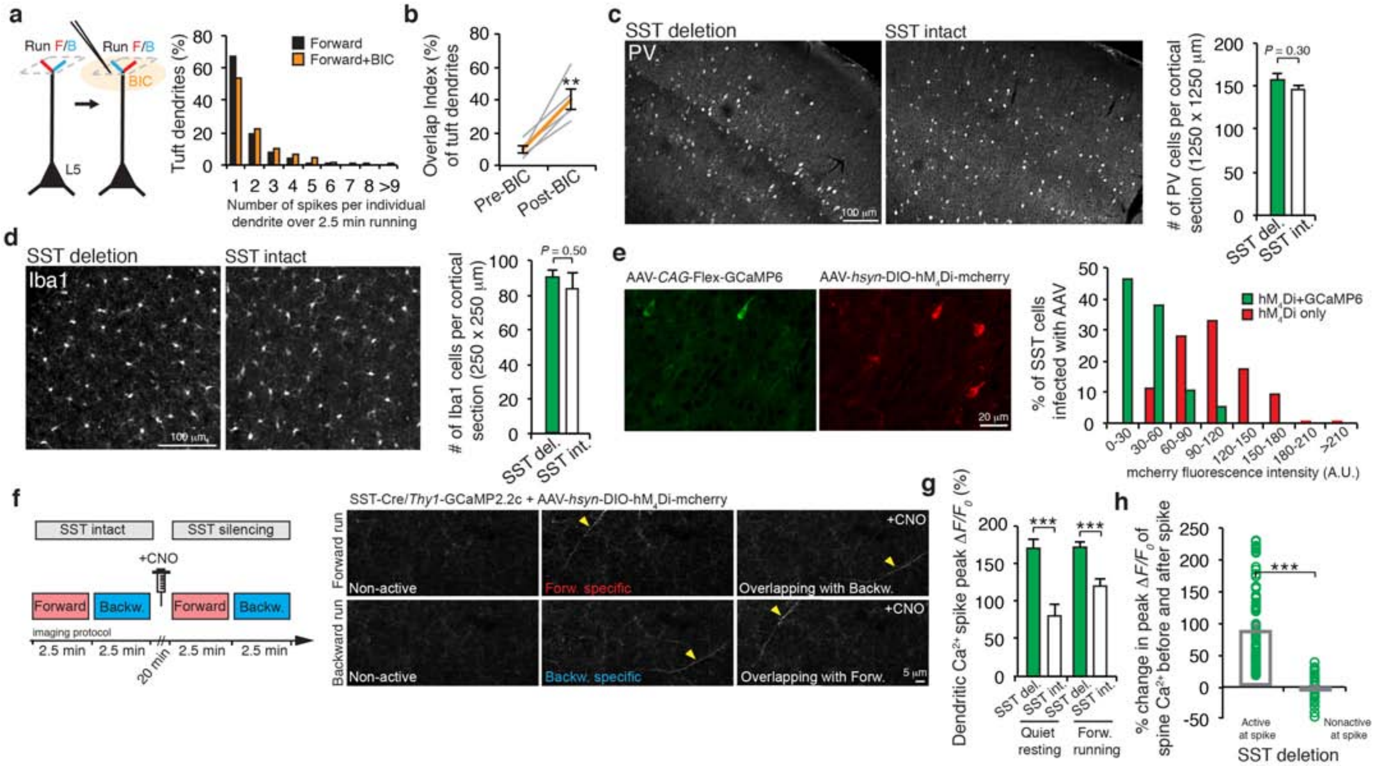
Extended Data Figure 5 | Ca^{2+} spikes cause long-lasting potentiation of task-related dendritic spines. **a**, Frequency distribution of the number of Ca^{2+} transients in individual spines detected by GCaMP6s over 2.5-min running ($n = 199$ spines from 14 dendritic branches). **b**, Frequency distribution of the peak amplitude ($\Delta F/F_0$) of spine Ca^{2+} transients on apical tuft branches expressing GCaMP6s during running. **c**, Frequency distribution of spine Ca^{2+} transient duration during running. **d**, Frequency distribution of the number of spine Ca^{2+} transients detected by GCaMP2.2c over 2.5-min running. **e**, Frequency distribution of the peak amplitude ($\Delta F/F_0$) of spine Ca^{2+} transients induced by forward running in tuft dendrites expressing GCaMP2.2c. As expected, spine Ca^{2+} transients detected with GCaMP6s were significantly higher than those detected by GCaMP2.2c in terms of amplitude (>200%) and frequency (>150%). **f**, Task-specific activation of dendritic spines detected by GCaMP2.2c during forward and backward running ($n = 98$ spines from 12 dendrites). **g**, Two-photon images of an apical tuft dendrite expressing GCaMP2.2c. An active spine (green) denoted by yellow arrowhead and a Ca^{2+} spike (red) are shown. **h**, **i**, Representative two-photon images and fluorescent traces of potentiated spines on L5 apical tuft dendritic segments expressing GCaMP2.2c and 5G. Ca^{2+} spikes occurred during trial 5 in **h** and **i**. **j**, Many spines were active before, during and after the generation of Ca^{2+} spikes. Horizontal lines indicate the start and end of spine Ca^{2+} transients. Variable Ca^{2+} spike duration indicated by red shaded bar. **k**, **l**, Comparison of spine head and shaft fluorescence for spines active at the time of spike generation versus neighbouring spines that are not active at the time of spike generation. $\Delta F/F_0$ of active spine heads are significantly larger ($P < 0.001$; 236.5% (GCaMP6s); 319.2% (GCaMP2.2c)) than that of neighbouring inactive spines during Ca^{2+} spike generation. **m**, Percentage change in average peak $\Delta F/F_0$ of spine Ca^{2+} transients before and after spikes detected by GCaMP6s (green) and GCaMP2.2c (red). The peaks of 3–4 spine Ca^{2+} transients ($\Delta F/F_0$) were averaged before and after the spike. There was no significant difference in the degree of spine Ca^{2+} transient potentiation detected by GCaMP6s (green) and GcaMP2.2c (red) ($P = 0.22$, unpaired *t*-test). **n**, Percentage change in the peak $\Delta F/F_0$ of individual spine Ca^{2+} transients (no average) immediately before and after spikes detected by GCaMP6s (green) and GCaMP2.2c (red) ($P = 0.16$, unpaired *t*-test). **o**, Comparison of spine size versus the percentage change in the peak amplitude of spine Ca^{2+} transients at 2 min and 40 min post spike. Data are mean ± s.e.m. Scale bars, 5 μm.



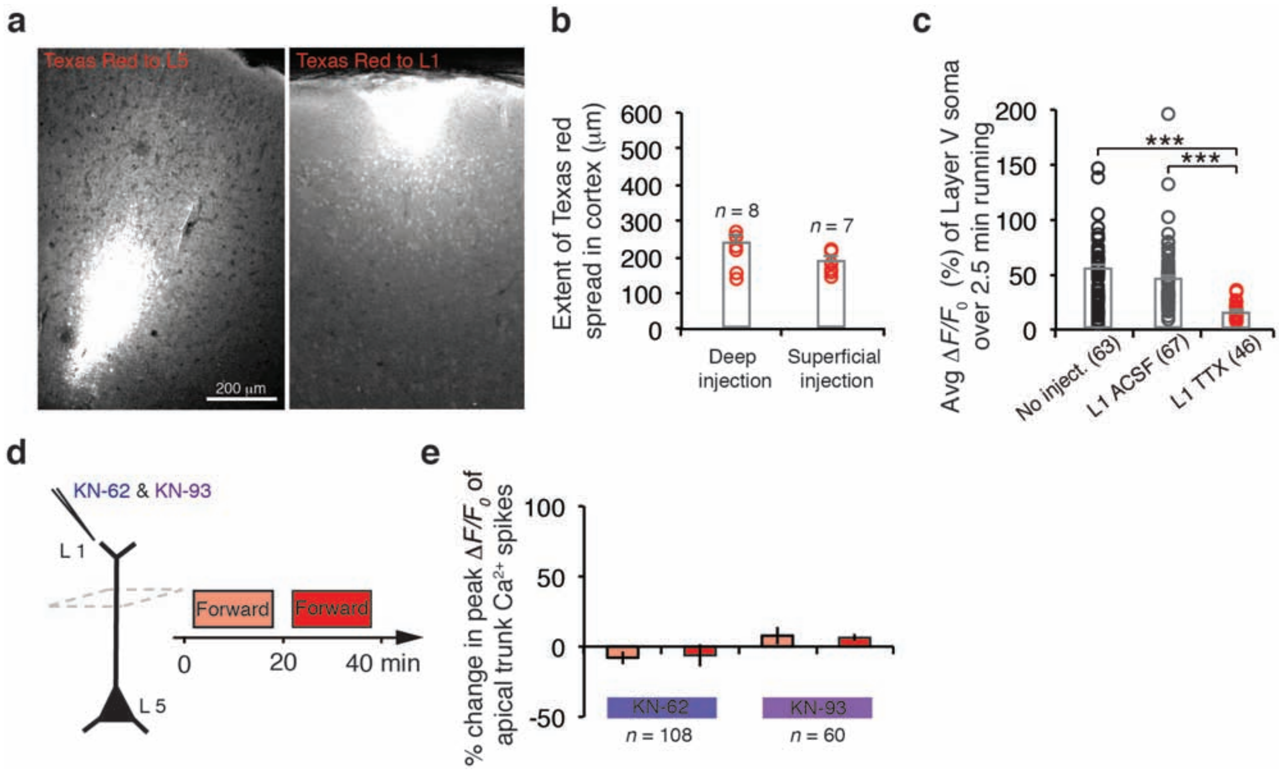
Extended Data Figure 6 | CaMKII inhibitors block potentiation of task-related dendritic spines but not Ca^{2+} spike generation. **a**, There is no correlation between the percentage change in the peak amplitude of spine Ca^{2+} transients and the peak amplitude of dendritic Ca^{2+} spike during forward running in the presence of CaMKII inhibitors (KN-62: $P = 0.08$; KN-93: $P = 0.42$, Pearson's correlation). **b**, Two-photon images of apical tuft dendrites expressing GCaMP2.2c during quiet resting state and forward running in the presence of ACSF, KN-93 and KN-62. **c**, **d**, Local application of CaMKII inhibitors, KN-93 and KN-62, to layer 1 did not induce significant changes in Ca^{2+} spike frequency or peak amplitude as compared to ACSF controls ($P > 0.05$, unpaired t-test). Scale bar, 10 μm. Data are mean ± s.e.m.



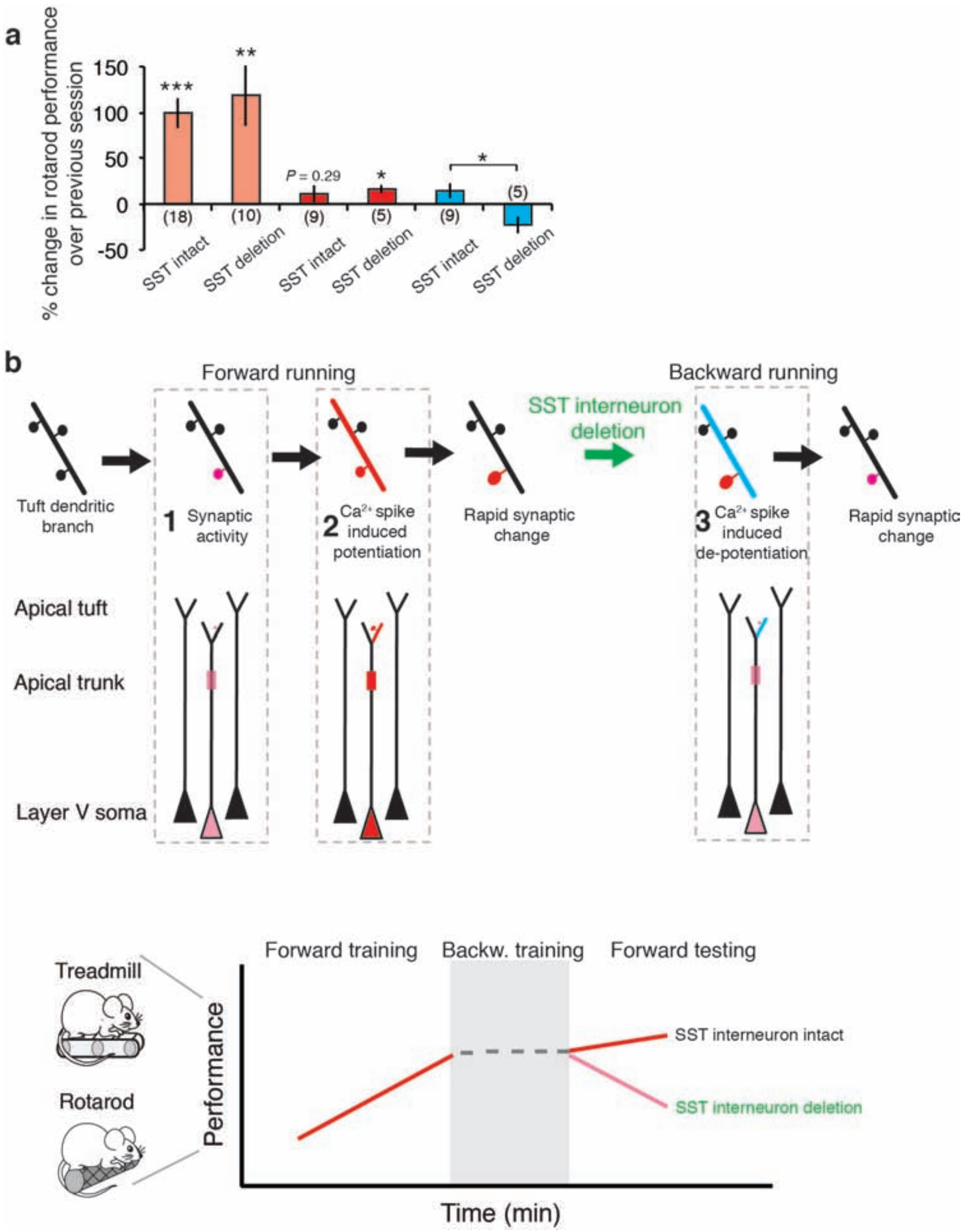
Extended Data Figure 7 | Spine activity relative to spike generation and task switching. **a**, **b**, In mice running forward, most spines (54%, red) exhibited activities at the time of the spike generation (near $t = 0$). Only a small fraction of spines were active <5 s before Ca^{2+} spike generation (17%, green). Thus, most spines active during forward running coincide with the spike generation. Of the spines that were active asynchronously relative to Ca^{2+} spikes, the average time interval between the two events was 12.2 ± 1.6 s. **c**, **d**, Changes in the peak amplitude of Ca^{2+} transients in forward running-activated spines versus the time interval between spine activity and the onset of task switching from forward to backward running. **e**, Spines active <5 s or >5 s before task switching (no backward spike) show no significant reduction in the peak amplitude of Ca^{2+} transients afterwards (<5 s: $P = 0.31$; $n = 9$; >5 s: $P = 0.84$; $n = 22$; unpaired t-test). **f**, There are no significant changes in the peak amplitude of spine Ca^{2+} transients (no backward spike) during the first and second half of a 30-s forward running trial, either before or after the task switching (paired t-test). Thus, spine transients do not gradually decrease during the initial forward running and recover to their highest values after the second forward running session is switched on. This suggests that the depotentation of spines active within 5 s before spikes in Fig. 3 is related to their interactions with local Ca^{2+} spikes.



Extended Data Figure 8 | The effect of inactivating SST neurons on Ca^{2+} spikes and potentiation of task-related dendritic spines. **a**, Experimental design for perturbing branch-specific Ca^{2+} spikes by applying bicuculline (BIC) to L1. Local bicuculline administration did not have a significant effect ($P > 0.05$, paired *t*-test) on the number of Ca^{2+} spikes generated during forward running as compared to control mice. Before bicuculline: 60.2 ± 10.2 Ca^{2+} spikes over 2.5 min; after bicuculline application: 77.4 ± 13.7 Ca^{2+} spikes over 2.5 min. More than 200 Ca^{2+} spikes were measured under each condition from five mice. **b**, Local bicuculline administration increased the percentage of apical tuft branches exhibiting Ca^{2+} spikes during both forward and backward running ($n = 5$ mice, $P = 0.007$, paired *t*-test). **c**, Coronal sections of the motor cortex from control or SST-deleted mice stained for parvalbumin (PV) 2 days after diphtheria toxin administration. No significant effect on the number of parvalbumin cells was observed ($P = 0.30$, unpaired *t*-test). **d**, Coronal sections of motor cortex from control or SST-deleted mice stained for microglia (Iba1) 2 days after diphtheria toxin administration. No significant effect on the number of microglia was observed ($P = 0.50$, unpaired *t*-test). **e**, Two-photon images and quantification of SST cells expressing GCaMP6 and hM_4Di-mCherry. **f**, Training protocol (left) to test the effect of inactivating SST cells after CNO treatment on branch-specific Ca^{2+} spike generation. Two-photon images (right) of apical tuft dendrites expressing GCaMP2.2c during forward and backward running before and after SST neuron silencing. Note that tuft dendrites after CNO exhibited Ca^{2+} spikes during both forward and backward running. **g**, The average peak amplitude of Ca^{2+} spikes during resting and running in SST-deleted and control mice expressing GCaMP2.2c. The peak $\Delta F/F_0$ of Ca^{2+} spikes during quiet resting and forward running in SST-deleted mice was significantly higher than in control mice ($P < 0.001$, unpaired *t*-test). **h**, In SST-deleted mice, spines active at the time of spike generation show enhanced Ca^{2+} signals after spike generation ($n = 49$. $P < 0.001$, paired *t*-test)), whereas spines not active at the spike generation ($n = 27$ spines) showed no significant increase in spine Ca^{2+} signals. In SST-deleted mice, the ratio of Ca^{2+} fluorescence intensity between spine heads and neighbouring shafts was 2.03 ± 0.23, significantly higher than that for neighbouring inactive spines (0.61 ± 0.03). Data are mean ± s.e.m. $^{**}P < 0.01$, $^{***}P < 0.001$.



Extended Data Figure 9 | CaMKII inhibitors block training-related potentiation of Ca^{2+} transients in L5 apical trunk nexus. **a**, Texas Red dye injection in L5 or L1. **b**, Fluorescence signals spread to a region with a diameter ~237 ± 19 μm in L5 and 188 ± 10 μm in L1 respectively. There is no overlap between the sites of injection in L5 and L1. **c**, Average $\Delta F/F_0$ of L5 soma detected by GCaMP6s over 2.5 min running under various conditions (no injection, L1 ACSF injection, L1 TTX injection) ($P < 0.001$, unpaired t-test). **d**, **e**, In control mice, local application of CaMKII inhibitors (KN-62 and KN-93) blocked the increase in Ca^{2+} transients at the apical trunk nexus after an initial and second training session of forward running.



Extended Data Figure 10 | Deletion of somatostatin-expressing interneuron impairs performance improvement after motor learning. **a**, Rotarod performance in control and SST-deleted mice subjected to forward–forward running or forward–backward running. When forward running on the accelerated rotarod was followed by backward running, SST-deleted mice displayed a reduction in their performance (the average speed animals achieved) as compared to control mice when tested again in forward running ($P < 0.05$, unpaired t-test). Data are mean ± s.e.m. **b**, A model showing the importance of BSDCS for inducing synaptic changes that affect L5 neuronal output during motor skill learning. Dendritic spines active (1, pink) at the time of spike generation (2) show enhanced Ca^{2+} activity and changes in synaptic strength (red) following the Ca^{2+} spike. Ca^{2+} spike-induced potentiation of synapses contributes to persistent synaptic changes, potentiated Ca^{2+} activity at the apical trunk and L5 soma, as well as improvements in performance over training sessions. In control mice, different motor tasks (that is, forward, backward running) induce Ca^{2+} spikes on different tuft branches of L5 neurons (not shown). Inactivation of SST interneurons results in individual dendritic branches generating Ca^{2+} spikes in response to both tasks (3). Loss of spatial segregation of Ca^{2+} spikes results in the depotentiation of synaptic changes-induced by previous learning in tuft dendritic branches and reduced Ca^{2+} activity in the apical trunk nexus and L5 soma when a different task is learned. SST interneuron inactivation induces a state of interference that impairs motor performance when several tasks are learned.

ARTICLE

doi:10.1038/nature14299

In vivo genome editing using *Staphylococcus aureus* Cas9

F. Ann Ran[1,2*], Le Cong[1,3*], Winston X. Yan[1,4,5*], David A. Scott[1,6,7], Jonathan S. Gootenberg[1,8], Andrea J. Kriz[3], Bernd Zetsche[1], Ophir Shalem[1], Xuebing Wu[9,10], Kira S. Makarova[11], Eugene V. Koonin[11], Phillip A. Sharp[3,9] & Feng Zhang[1,6,7,12]

The RNA-guided endonuclease Cas9 has emerged as a versatile genome-editing platform. However, the size of the commonly used Cas9 from *Streptococcus pyogenes* (SpCas9) limits its utility for basic research and therapeutic applications that use the highly versatile adeno-associated virus (AAV) delivery vehicle. Here, we characterize six smaller Cas9 orthologues and show that Cas9 from *Staphylococcus aureus* (SaCas9) can edit the genome with efficiencies similar to those of SpCas9, while being more than 1 kilobase shorter. We packaged SaCas9 and its single guide RNA expression cassette into a single AAV vector and targeted the cholesterol regulatory gene *Pcsk9* in the mouse liver. Within one week of injection, we observed >40% gene modification, accompanied by significant reductions in serum Pcsk9 and total cholesterol levels. We further assess the genome-wide targeting specificity of SaCas9 and SpCas9 using BLESS, and demonstrate that SaCas9-mediated *in vivo* genome editing has the potential to be efficient and specific.

Cas9, an RNA-guided endonuclease derived from the type II CRISPR-Cas bacterial adaptive immune system[1–7], has been harnessed for genome editing[8,9] and holds tremendous promise for biomedical research. Genome editing of somatic tissue in postnatal animals, however, has been limited in part by the challenge of delivering Cas9 *in vivo*. For this purpose, adeno-associated virus (AAV) vectors are attractive vehicles[10] because of their low immunogenic potential, reduced oncogenic risk from host-genome integration[11], and broad range of serotype specificity[12–15]. Nevertheless, the restrictive cargo size (~4.5 kb, excluding the inverted terminal repeats) of AAV presents an obstacle for packaging the commonly used *Streptococcus pyogenes* Cas9 (SpCas9, ~4.2 kb) and its single guide RNA (sgRNA) in a single vector; although technically feasible[17], this approach leaves little room for customized expression and control elements[16].

In search of smaller Cas9 enzymes for efficient *in vivo* delivery by AAV, we have previously described a short Cas9 from the CRISPR1 locus of *Streptococcus thermophilus* LMD-9 (St1Cas9, ~3.3 kb)[8] as well as a rationally-designed truncated form of SpCas9 (ref. 18) for genome editing in human cells. However, both systems have important practical drawbacks: the former requires a complex protospacer-associated motif (PAM) sequence (NNAGAAW)[3], which restricts the range of accessible targets, whereas the latter exhibits reduced activity. Given the substantial diversity of CRISPR-Cas systems present in sequenced microbial genomes[19], we therefore sought to interrogate and discover additional Cas9 enzymes that are small, efficient and broadly targeting.

In vitro cleavage by small Cas9 enzymes

Type II CRISPR-Cas systems require only two main components for eukaryotic genome editing: a Cas9 enzyme, and a chimaeric sgRNA[6] derived from the CRISPR RNA (crRNA) and the noncoding trans-activating crRNA (tracrRNA)[4,20]. Analysis of over 600 Cas9 orthologues shows that these enzymes are clustered into two length groups with characteristic protein sizes of approximately 1,350 and 1,000 amino acid residues, respectively[19,21] (Extended Data Fig. 1a), with shorter Cas9 enzymes having significantly truncated REC domains (Fig. 1a). From these shorter Cas9 enzymes, which belong to Type IIA and IIC subtypes, we selected six candidates for profiling (Fig. 1a and Extended Data Fig. 1b). To determine the cognate crRNA and tracrRNA for each Cas9, we computationally identified regularly interspaced repeat sequences (direct repeats) within a 2-kb window flanking the CRISPR locus. We then predicted the tracrRNA by detecting sequences with strong complementarity to the direct repeat sequence (an anti-repeat region), at least two predicted stem-loop structures, and a Rho-independent transcriptional termination signal up to 150 nucleotides downstream of the anti-repeat region. Although a truncated tracrRNA can support robust DNA cleavage *in vitro*[6], previous reports show that the secondary structures of the tracrRNA are important for Cas9 activity in mammalian cells[8,9,18,22]. Therefore, we designed sgRNA scaffolds for each orthologue by fusing the 3′ end of a truncated direct repeat with the 5′ end of the corresponding tracrRNA, including the full-length tail, via a 4-nucleotide linker[6] (Extended Data Fig. 1b and Supplementary Table 1). To identify the PAM sequence for each Cas9, we first constructed a library of plasmid DNA containing a constant 20-bp target followed by a degenerate 7-bp sequence (5′-NNNNNNN). We then incubated cell lysate from human embryonic kidney 293FT (293FT) cells expressing the Cas9 orthologue with its *in vitro*-transcribed sgRNA and the plasmid library. By generating a consensus from the 7-bp sequence found on successfully cleaved DNA plasmids (Fig. 1b), we determined putative PAMs for each Cas9 (Fig. 1c). We observed that, similar to SpCas9, most Cas9 orthologues cleaved targets 3-bp upstream of the PAM (Extended Data Fig. 2). To validate each putative PAM from the library, we then incubated a DNA template bearing the consensus PAM with cell lysate

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [2]Society of Fellows, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Graduate Program in Biophysics, Harvard Medical School, Boston, Massachusetts 02115, USA. [5]Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School, Boston, Massachusetts 02115, USA. [6]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [7]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [8]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [9]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [10]Computational and Systems Biology Graduate Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [11]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA. [12]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.

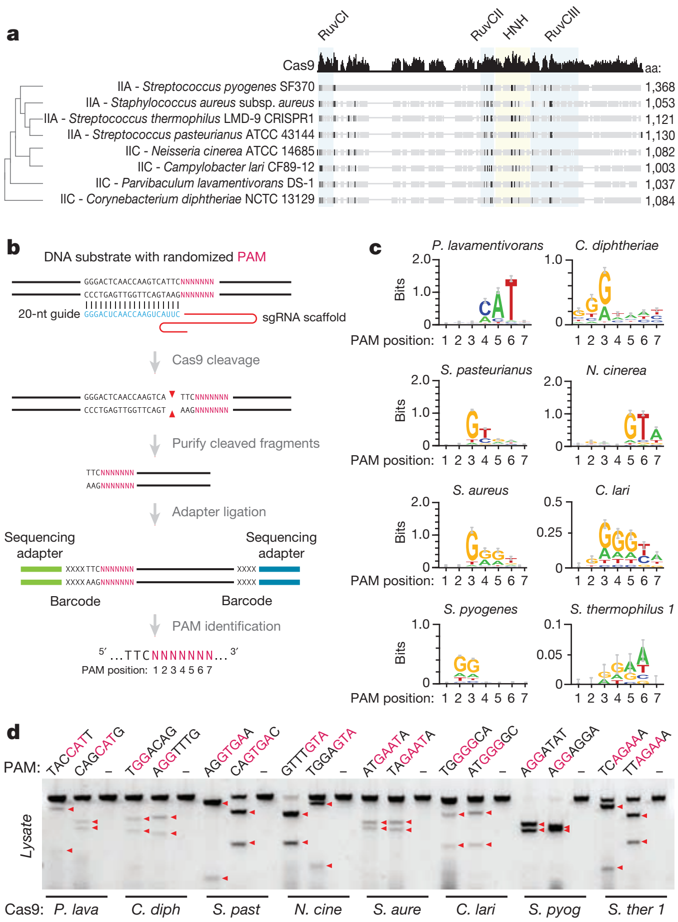


Figure 1 | Biochemical screen for small Cas9 orthologues. a, Phylogenetic tree of selected Cas9 orthologues. Subfamily and sizes (amino acids) are indicated, with nuclease domains highlighted in coloured boxes, and conserved sequences in black. **b**, Schematic illustration of the *in vitro* cleavage-based method used to identify the first seven positions (5′-NNNNNNN) of protospacer adjacent motifs (PAMs). **c**, Consensus PAMs for eight Cas9 orthologues from sequencing of cleaved fragments. Error bars are Bayesian 95% confidence interval[45]. **d**, Cleavage using different orthologues and sgRNAs targeting loci bearing the putative PAMs (consensus shown in red). Red triangles indicate cleavage fragments.

and the corresponding sgRNA. We found that the Cas9 orthologues, in combination with the sgRNA designs, successfully cleaved the appropriate targets (Fig. 1d and Supplementary Table 2).

To test whether each Cas9 orthologue can facilitate genome editing in mammalian cells, we co-transfected 293FT cells with individual Cas9 enzymes and their respective sgRNAs targeting human endogenous loci containing the appropriate PAMs. Of the six Cas9 orthologues tested, only the one from *Staphylococcus aureus* (SaCas9) produced indels with efficiencies comparable to those of SpCas9 (Extended Data Fig. 3a, b and Supplementary Table 3), suggesting that DNA-cleavage activity in cell-free assays does not necessarily predict activity in mammalian cells. These observations prompted us to focus on harnessing SaCas9 and its sgRNA for *in vivo* applications.

SaCas9 sgRNA design and PAM discovery

Although mature crRNAs in *S. pyogenes* are processed to contain 20-nucleotide spacers (guides) and 19- to 22-nucleotide direct repeats[4], RNA sequencing of crRNAs from other organisms reveals that the spacer and direct repeat sequence lengths can vary[4,20,23]. We therefore tested sgRNAs for SaCas9 with variable guide lengths and repeat:anti-repeat duplexes. We found that SaCas9 achieves the highest editing efficiency in mammalian cells with guides between 21 and 23 nucleotides long and can accommodate a range of lengths for the direct repeat:anti-repeat region (Fig. 2a, b, Extended Data Fig. 4). This notably contrasts with SpCas9, where the natural 20-nucleotide guide length can be truncated to 17 nucleotides without significantly compromising nuclease activity, while increasing specificity[24]. Additionally, replacing the first base of the guide with guanine further improved SaCas9 activity (Extended Data Fig. 3c).

To fully characterize the SaCas9 PAM and the seed region within its guide sequence[25], we performed chromatin immunoprecipitation (ChIP) using catalytically mutant forms of SaCas9 (dSaCas9, D10A and N580A mutations, based on homology to SpCas9) or SpCas9 (dSpCas9, D10A and H840A mutations) and their corresponding sgRNAs. We targeted two loci in the human *EMX1* gene with composite NGGRRT PAMs, which allow targeting by both dCas9s. A search for motifs containing both the guide region and PAM within 50 nucleotides of the ChIP peak summits revealed seed sequences of 7–8 nucleotides for dSaCas9 (Fig. 2c). In addition, NNGRRT and NGG PAMs were found adjacent to the seed sequences for dSaCas9 and dSpCas9, respectively (Extended Data Fig. 5). Although the 6th position of the PAM is predominantly thymine, we did observe low levels of degeneracy in both the biochemical and ChIP-based PAM discovery assays (Fig. 1c and Extended Data Fig. 5a). We therefore tested the base preference for this position and determined that, although SaCas9 cleaves genomic targets most efficiently with NNGRRT, all NNGRR PAMs can be cleaved and should be considered as potential targets, especially in the context of off-target evaluations (Fig. 2d, Extended Data Fig. 6 and Supplementary Table 4).

Unbiased profiling of Cas9 specificity

As advances in Cas9 technology promise to enable a broad range of *in vivo* and therapeutic applications, accurate, genome-wide identification of off-target nuclease activity has become increasingly important. Although a number of studies have employed sequence similarity-based off-target search[22,26–30] or dCas9-ChIP[31,32] to predict off-target sites for Cas9, such approaches cannot assess the nuclease activity of Cas9 in a comprehensive and unbiased manner. To measure the genome-wide

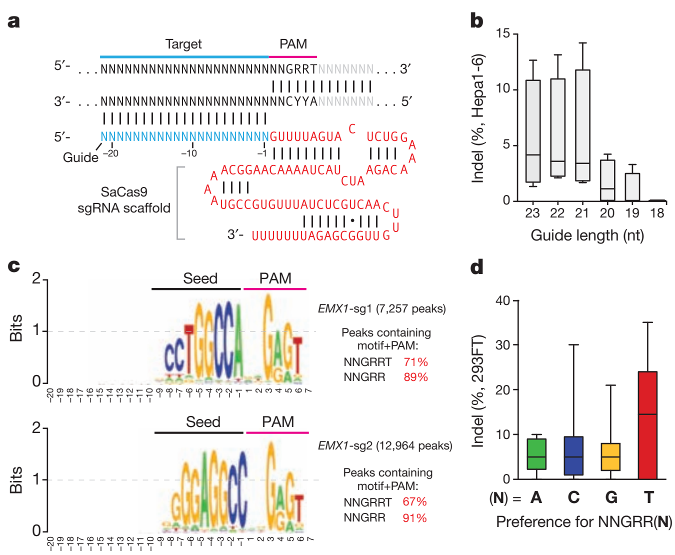


Figure 2 | Characterization of *Staphylococcus aureus* Cas9 (SaCas9) in 293FT cells. a, SaCas9 sgRNA scaffold (red) and guide (blue) base-pairing at target locus (black) immediately 5′ of PAM. **b**, Box-whisker plot showing indel levels vary depending on the length of the guide sequence ($n = 4$). **c**, dSaCas9-ChIP reveals peaks associated with seed + PAM. Text to the right indicates the total number of peaks and percentage containing significant (false discovery rate < 0.1) match to the guide motif followed by NNGRRT or NNGRR PAMs. **d**, Pooled indel values for NNGRR(A), (C), (G), or (T) PAM combinations ($n = 12$, 21, 39 and 44, respectively).

cleavage activity of SaCas9 and SpCas9 directly, we applied BLESS (direct *in situ* breaks labelling, enrichment on streptavidin and next-generation sequencing)[33] to capture a snapshot of Cas9-induced DNA double-stranded breaks (DSBs) in cells. We transfected 293FT cells with SaCas9 or SpCas9 and the same *EMX1* targeting guides used in the previous ChIP experiment, or pUC19 as a negative control. After cells are fixed, free genomic DNA ends from DSBs are captured using biotinylated adaptors and analysed by deep sequencing (Fig. 3a). To identify candidate Cas9-induced DSB sites genome-wide, we established a three-step analysis pipeline following alignment of the sequenced BLESS reads to the genome (Extended Data Fig. 7a, Supplementary Discussion). First, we applied nearest-neighbour clustering on the aligned reads to identify groups of DSBs (DSB clusters) across the genome. Second, we sought to separate potential Cas9-induced DSB clusters from background DSB clusters resulting from low frequency biological processes and technical artefacts, as well as high-frequency telomeric and centromeric DSB hotspots[33]. From the on-target and a small subset of verified off-target sites (predicted by sequence similarity using a previously established method[22] and sequenced to detect indels), we found that reads in Cas9-induced DSB clusters mapped to characteristic, well-defined genomic positions compared to the more diffuse alignment pattern at background DSB clusters. To distinguish between the two types of DSB clusters, we calculated in each cluster the distance between all possible pairs of forward and reverse-oriented reads (corresponding to 3′ and 5′ ends of DSBs), and filtered out the background DSB clusters based on the distinctive pairwise-distance distribution of these clusters (Extended Data Fig. 7b, c). Third, the DSB score for a given locus was calculated by comparing the count of DSBs in the experimental and negative control samples using a maximum-likelihood estimate[22] (Supplementary Discussion). This analysis identified the on-target loci for both SaCas9 and SpCas9 guides as the top scoring sites, and revealed additional sites with high DSB scores (Fig. 3b–d).

Next, we sought to assess whether DSB scores correlated with indel formation. We used targeted deep sequencing to detect indel formation

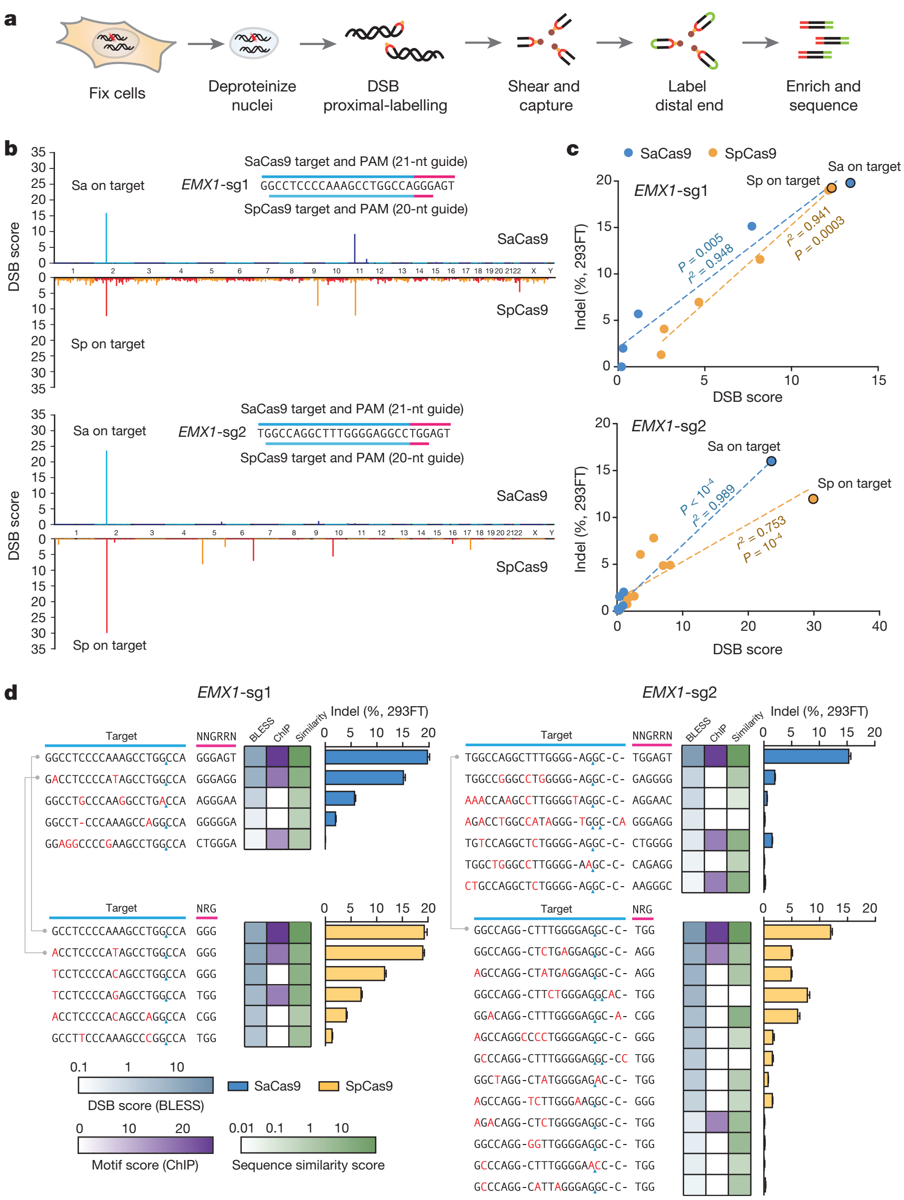


Figure 3 | Characterization of genome-wide nuclease activity of SaCas9 and SpCas9. **a**, Schematic of BLESS processing steps. **b**, Manhattan plots of genome-wide DSB clusters generated by each Cas9 and sgRNA pair, with on-target loci shown above (see Supplementary Discussion). **c**, Correlation between DSB scores and indel levels for top-scoring DSB clusters. Trend lines, r^2 and P values are calculated using ordinary least squares method. **d**, Off-target loci from BLESS with detectable indels through targeted deep sequencing ($n = 3$) are shown. Heat maps indicate DSB score (blue), motif score from ChIP (purple), or sequence similarity score (green) for each locus. Blue triangles indicate peak positions of BLESS signal.

on the ~30 top-ranking off-target sites identified by BLESS for each Cas9 and sgRNA combination. We found that only those sites that contained PAM and homology to the guide sequence exhibited indels (Extended Data Fig. 8). We observed a strong linear correlation between DSB scores and indel levels for each Cas9 and sgRNA pairing ($r^2 = 0.948$ and 0.989 for the two *EMX1* targets with SaCas9 and $r^2 = 0.941$ and 0.753 for those with SpCas9) (Fig. 3c, Extended Data Fig. 9b–d). Furthermore, BLESS identified additional off-target sites not previously predicted by sequence similarity to target or ChIP (Extended Data Figs 7 and 9, Supplementary Tables 5 and 6). These new off-target sites include not only those containing Watson–Crick base-pairing mismatches to the guide, but also the recently reported insertion and deletion mismatches in the guide:target heteroduplex (Fig. 3d)[29,30]. Together, these results highlight the need for more precise understanding of rules governing Cas9 nuclease activity, a requisite step towards improving the predictive power of computational guide design programs.

In vivo genome editing using SaCas9

Following *in vitro* characterization, we incorporated SaCas9 and its sgRNA into an AAV vector to test its efficacy and specificity *in vivo*. The small size of SaCas9 enables packaging of both a U6-driven sgRNA and a cytomegalovirus (CMV)- or thyroxine-binding globulin (TBG)-driven SaCas9 expression cassette into a single AAV vector within the 4.5-kb packaging limit. Using hepatocyte-tropic AAV serotype 8, we targeted the mouse apolipoprotein (*Apob*) gene (Extended Data Fig. 10a). One week after intravenous administration of virus into C57BL/6 mice, we observed ~5% indel formation in liver tissue; after four weeks, the liver tissue showed characteristic hepatic lipid accumulation from Apob knockdown following histology analysis using oil red staining[34–37] (Extended Data Fig. 10b, c).

We next targeted proprotein convertase subtilisin/kexin type 9 (*Pcsk9*), a therapeutically relevant gene involved in cholesterol homeostasis[38]. Inhibitors of the human convertase PCSK9 have emerged as a promising new class of cardioprotective drugs, after human genetic studies revealed that loss of *PCSK9* is associated with a reduced risk of cardiovascular disease and lower levels of low-density lipoprotein (LDL) cholesterol[39–41]. We designed two *Pcsk9*-targeting sgRNAs (20-nucleotide guides with additional 5′ guanine) and validated their activity *in vitro*. Each sgRNA was packaged into AAV-SaCas9 and injected into mice (2×10^{11} total genome copies) (Fig. 4a). One week after administration, we observed greater than 40% indel formation at either locus in whole liver tissue, with similar levels two and four weeks post-injection (Fig. 4b). To determine the effect of *Pcsk9*-targeting AAV-SaCas9 dosage on serum Pcsk9 and total cholesterol levels, we administered a range of AAV titres from 0.5×10^{11} to 4×10^{11} total genome copies. With all titres, we observed a ~95% decrease in serum Pcsk9 and a ~40% decrease in total cholesterol one week after administration, both of which were sustained throughout the course of four weeks (Fig. 4c, d).

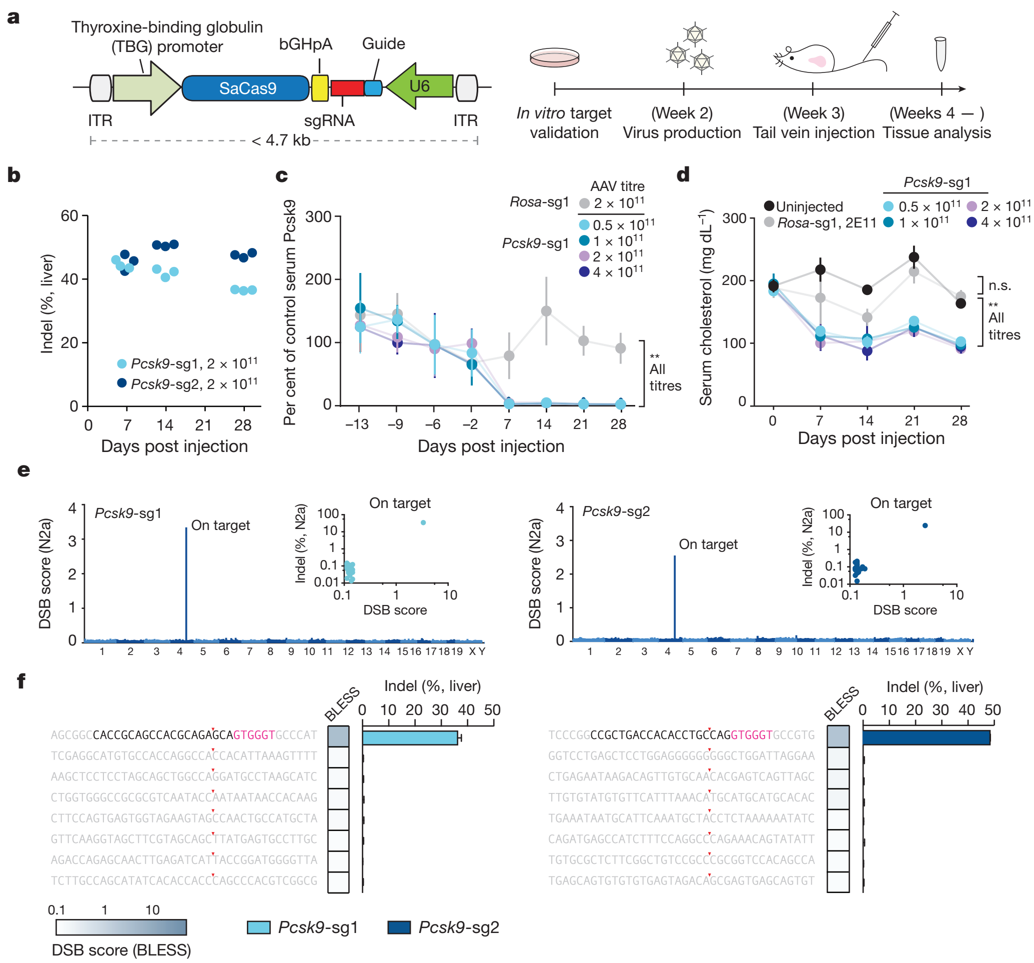


Figure 4 | AAV-delivery of SaCas9 for *in vivo* genome editing. **a**, Single-vector AAV system and experimental timeline. **b**, Indels at *Pcsk9* targets in liver tissue following injection of AAV at 2×10^{11} total genome copies ($n = 3$ animals). **c**, **d**, Time course of serum Pcsk9 (**c**) and total cholesterol in animals (**d**; $n = 3$ for all titres and time points, error bars show s.e.m.). **e**, Manhattan plots of BLESS-identified DSB clusters in N2a cells. Inset indicates indel levels at top DSB scoring loci. **f**, Indels in liver tissue ($n = 3$ animals, error bars indicate Wilson intervals) at BLESS-identified off-target loci from N2a cells. Heat map indicates DSB scores.

We next considered SaCas9 off-target modifications in the liver tissue samples. To first identify candidate off-target cleavage sites for the two *Pcsk9*-targeting guides, we transiently transfected an AAV-CMV::SaCas9 vector into mouse Neuroblastoma-2a (N2a) cells and applied BLESS to detect Cas9-induced DSBs in the genome. For both guides, we found very low levels of DSB signal across the genome except at the on-target loci (Fig. 4e). Targeted deep sequencing of the candidate off-target sites identified by BLESS in N2a cells did not reveal appreciable levels of indels in either N2a cells or liver tissue (4 weeks post injection of 2×10^{11} total genome copies) (Fig. 4e, f and Supplementary Table 8). We additionally sequenced off-target sites predicted by target sequence similarity, and likewise did not detect indel formations (Supplementary Table 9).

Finally, we examined the titre-matched *Pcsk9*-targeting and enhanced green fluorescent protein-conjugated (EGFP) TBG–EGFP cohorts as well as naive animals for signs of toxicity or acute immune response. At 1 week post-injection, necropsy and gross examination of liver tissue of the cohorts revealed no abnormalities; further histological examination of the liver by haematoxylin and eosin (H&E) staining showed no signs of inflammation, such as aggregates of lymphocytes or macrophages (Fig. 5a). Throughout the time course of the experiment, there were no elevated levels of serum alanine aminotransferase (ALT), albumin, and total bilirubin in any of the cohorts. We observed a slight trend in aspartate transaminase (AST) increase across all cohorts at four weeks, including the uninjected animals. The elevated levels did not exceed the upper limit of normal and is not indicative of hepatocellular injury in animals (Fig. 5b). However, a larger cohort study should be conducted to further evaluate the potential side-effects of Cas9-mediated *in vivo* genome editing. In addition, the differences between mouse and human immune responses need to be better elucidated before considering this approach for therapeutic applications.

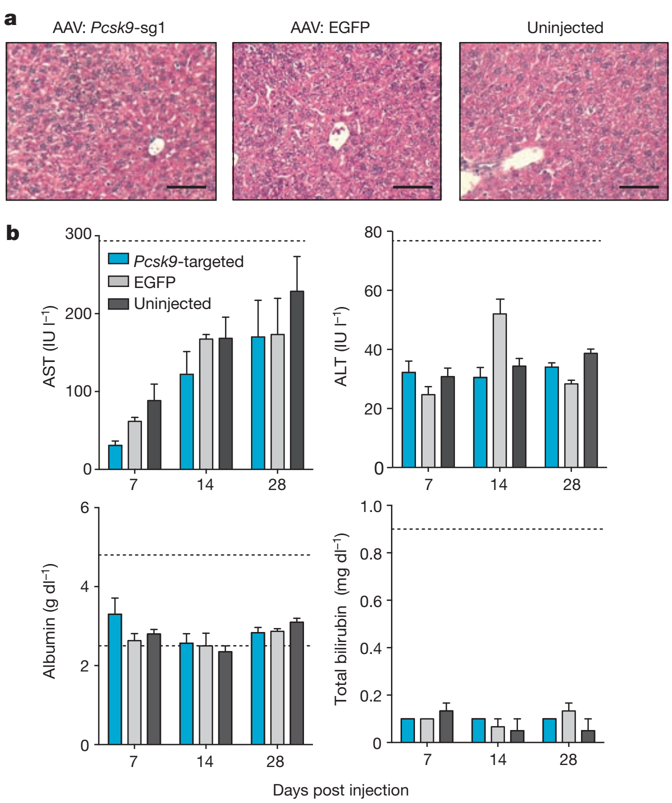


Figure 5 | Liver function tests and toxicity examination in injected animals. **a**, Histological analysis of the liver at 1 week post-injection by haematoxylin and eosin stain. Scale bars, 10 μm. **b**, Liver function tests in Pcsk9-targeted (both Pcsk9-sg1 and Pcsk9-sg2; 2×10^{11} total genome copies, $n \geq 4$), TBG–EGFP-injected (2×10^{11} total genome copies, $n = 3$), and uninjected ($n = 5$) animals. Dashed lines show the upper and lower ranges of normal value in mice where applicable.

Discussion

Here, we develop a small and efficient Cas9 from *S. aureus* for *in vivo* genome editing[17]. The results of these experiments highlight the power of using comparative genomic analysis[19,42] in expanding the CRISPR-Cas9 toolbox. Identification of new Cas9 orthologues[19,42], in addition to structure-guided engineering, could yield a repertoire of Cas9 variants with expanded capabilities and minimized molecular weight, for nucleic acid manipulation to further advance genome and epigenome engineering.

The AAV-SaCas9 system is able to mediate efficient and rapid editing of *Pcsk9* in the mouse liver, resulting in reductions of serum Pcsk9 and total cholesterol levels. To assess the specificity of SaCas9, we used an unbiased DSB detection method, BLESS, to identify a list of candidate off-target cleavage sites in a mouse cell line. We examined these sites in liver tissue transduced by AAV-SaCas9 and did not observe any indel formation within the detection limits of *in vitro* BLESS and targeted deep sequencing. Importantly, the off-target sites identified *in vitro* might differ from those *in vivo*, which need to be further evaluated by the applications of BLESS or other unbiased techniques such as those published during the revision of this work[43,44]. Finally, we did not observe any overt signs of acute toxicity in mice at one to four weeks after virus administration. Although more studies are needed to further improve the SaCas9 system for *in vivo* genome editing, such as assessing the long-term impact of Cas9 and sgRNA expression, these findings suggest that *in vivo* genome editing using SaCas9 has the potential to be highly efficient and specific.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 February 2014; accepted 5 February 2015.

Published online 1 April 2015.

1. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).
2. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
3. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).
4. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).
5. Sapranauskas, R. *et al.* The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39,** 9275–9282 (2011).
6. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
7. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109,** E2579–E2586 (2012).
8. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
9. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826 (2013).
10. Gaudet, D. *et al.* Review of the clinical development of alipogene tiparvovec gene therapy for lipoprotein lipase deficiency. *Atheroscler. Suppl.* **11,** 55–60 (2010).
11. Vasileva, A. & Jessberger, R. Precise hit: adeno-associated virus in gene targeting. *Nature Rev. Microbiol.* **3,** 837–847 (2005).
12. Mingozzi, F. & High, K. A. Therapeutic *in vivo* gene transfer for genetic disease using AAV: progress and challenges. *Nature Rev. Genet.* **12,** 341–355 (2011).
13. Gao, G., Vandenberghe, L. H. & Wilson, J. M. New recombinant serotypes of AAV vectors. *Curr. Gene Ther.* **5,** 285–297 (2005).
14. Kay, M. A. State-of-the-art gene-based therapies: the road ahead. *Nature Rev. Genet.* **12,** 316–328 (2011).
15. Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J. E. Analysis of AAV serotypes 1–9 mediated gene expression and tropism in mice after systemic injection. *Mol. Ther.* **16,** 1073–1080 (2008).
16. Swiech, L. *et al. In vivo* interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nature Biotechnol.* **33,** 102–106 (2015).
17. Senís, E. *et al.* CRISPR/Cas9-mediated genome engineering: an adeno-associated viral (AAV) vector toolbox. *Biotechnol. J.* **9,** 1402–1412 (2014).
18. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–949 (2014).

19. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42,** 6091–6105 (2014).
20. Chylinski, K., Le Rhun, A. & Charpentier, E. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.* **10,** 726–737 (2013).
21. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157,** 1262–1278 (2014).
22. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnol.* **31,** 827–832 (2013).
23. Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc. Natl Acad. Sci. USA* **110,** 15644–15649 (2013).
24. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnol.* **32,** 279–284 (2014).
25. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl Acad. Sci. USA* **108,** 10098–10103 (2011).
26. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnol.* **31,** 822–826 (2013).
27. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnol.* **31,** 833–838 (2013).
28. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnol.* **31,** 839–843 (2013).
29. Lin, Y. *et al.* CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.* **42,** 7473–7485 (2014).
30. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30,** 1473–1475 (2014).
31. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnol.* **32,** 670–676 (2014).
32. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature Biotechnol.* **32,** 677–683 (2014).
33. Crosetto, N. *et al.* Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nature Methods* **10,** 361–365 (2013).
34. Young, S. G. Recent progress in understanding apolipoprotein B. *Circulation* **82,** 1574–1594 (1990).
35. Soutschek, J. *et al.* Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* **432,** 173–178 (2004).
36. Rozema, D. B. *et al.* Dynamic PolyConjugates for targeted *in vivo* delivery of siRNA to hepatocytes. *Proc. Natl Acad. Sci. USA* **104,** 12982–12987 (2007).
37. Wolfrum, C. *et al.* Mechanisms and optimization of *in vivo* delivery of lipophilic siRNAs. *Nature Biotechnol.* **25,** 1149–1157 (2007).
38. Fitzgerald, K. *et al.* Effect of an RNA interference drug on the synthesis of proprotein convertase subtilisin/kexin type 9 (PCSK9) and the concentration of serum LDL cholesterol in healthy volunteers: a randomised, single-blind, placebo-controlled, phase 1 trial. *Lancet* **383,** 60–68 (2014).
39. Abifadel, M. *et al.* Mutations in *PCSK9* cause autosomal dominant hypercholesterolemia. *Nature Genet.* **34,** 154–156 (2003).
40. Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in *PCSK9*. *Nature Genet.* **37,** 161–165 (2005).
41. Horton, J. D., Cohen, J. C. & Hobbs, H. H. Molecular biology of PCSK9: its role in LDL metabolism. *Trends Biochem. Sci.* **32,** 71–77 (2007).
42. Briner, A. E. *et al.* Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol. Cell* **56,** 333–339 (2014).
43. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnol.* **33,** 187–197 (2015).
44. Frock, R. L. *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature Biotechnol.* **33,** 179–186 (2015).
45. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Charpentier, I. Fonfara and K. Chylinski for discussions; A. Scherer-Hoock, B. Clear and the MIT Division of Comparative Medicine for assistance with animal experiments; Boston Children's Hospital Viral Core and R. Xiao for assistance with AAV production; N. Crosetto for advice on BLESS; C.-Y. Lin and I. Slaymaker for experimental assistance; and the entire Zhang laboratory for support and advice. F.A.R. is a Junior Fellow at the Harvard Society of Fellows. W.X.Y. is supported by T32GM007753 from the National Institute of General Medical Sciences and a Paul and Daisy Soros Fellowship. J.S.G. is supported by a US Department of Energy Computational Science Graduate Fellowship. X.W. is a Howard Hughes Medical Institute International Student Research Fellow. P.A.S. is supported by United States Public Health Service grants RO1-GM34277, R01-CA133404 from the National Institutes of Health, and PO1-CA42063 from the National Cancer Institute, and partially by Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute. F.Z. is supported by the National Institutes of Health through NIMH (5DP1-MH100706) and NIDDK (5R01DK097768-03), a Waterman Award from the National Science Foundation, the Keck, New York Stem Cell, Damon Runyon, Searle Scholars, Merkin, and Vallee Foundations, and B. Metcalfe. F.Z. is a New York Stem Cell Foundation Robertson Investigator. The Children's Hospital virus core is supported by an NIH core grant (5P30EY012196-17). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. CRISPR reagents are available to the academic community through Addgene, and information about the protocols, plasmids, and reagents can be found at the Zhang laboratory website http://www.genome-engineering.org.

Author Contributions F.A.R. and F.Z. conceived this study. F.A.R., L.C., W.X.Y. and F.Z. designed and performed the experiments with help from all authors. F.A.R., J.S.G., O.S., K.S.M., E.V.K. and F.Z. contributed to analysis of Cas9 orthologues, crRNA and tracrRNA, and PAM. A.J.K., F.A.R., X.W., and P.A.S. led ChIP and computational analysis and validation. F.A.R., W.X.Y. and L.C. performed BLESS and targeted sequencing of BLESS-identified off-target sites, and D.A.S. contributed computational analysis of BLESS data. W.X.Y., F.A.R., L.C. and B.Z. contributed animal data. W.X.Y., F.A.R., L.C., J.S.G., and F.Z. wrote the manuscript with help from all authors.

Author Information All reagents described in this manuscript have been deposited with Addgene (plasmid IDs 61591, 61592 and 61593). Source data are available online and deep sequencing data are available at Sequence Read Archive under BioProject accession number PRJNA274149. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.Z. (zhang@broadinstitute.org).

METHODS

No statistical methods were used to predetermine sample size.

***In vitro* transcription and cleavage assay.** Cas9 orthologues were human codon-optimized and synthesized by GenScript, and transfected into 293FT cells as described below. Whole-cell lysates from 293FT cells were prepared with lysis buffer (20 mM HEPES, 100 mM KCl, 5 mM MgCl2, 1 mM DTT, 5% glycerol, 0.1% Triton X-100) supplemented with Protease Inhibitor Cocktail (Roche). T7-driven sgRNA was transcribed *in vitro* using custom oligonucleotides (Supplementary Information) and HiScribe T7 *In vitro* Transcription Kit (NEB), following the manufacturer's recommended protocol. The *in vitro* cleavage assay was carried out as follows: for a 20 μl cleavage reaction, 10 μl of cell lysate was incubated with 2 μl cleavage buffer (100 mM HEPES, 500 mM KCl, 25 mM $MgCl_2$, 5 mM DTT, 25% glycerol), 1 μg *in vitro* transcribed RNA and 200 ng EcoRI-linearized pUC19 plasmid DNA or 200 ng purified PCR amplicons from mammalian genomic DNA containing target sequence. After 30 min incubation, cleavage reactions were purified using QIAquick Spin Columns and treated with RNase A at final concentration of 80 ng μl^{-1} for 30 min and analysed on a 1% agarose E-Gel (Life Technologies).

***In vitro* PAM screen.** Rho-independent transcriptional termination was predicted using the ARNold terminator search tool[46,47]. For the PAM library, a degenerate 7-bp sequence was cloned into a pUC19 vector. For each orthologue, the *in vitro* cleavage assay was carried out as above with 1 μg T7-transcribed sgRNA and 400 ng pUC19 with degenerate PAM. Cleaved plasmids were linearized by NheI, gel extracted, and ligated with Illumina sequencing adaptors. Barcoded and purified DNA libraries were quantified by Quant-iT PicoGreen dsDNA Assay Kit or Qubit 2.0 Fluorometer (Life Technologies) and pooled in an equimolar ratio for sequencing using the Illumina MiSeq Personal Sequencer (Life Technologies). MiSeq reads were filtered by requiring an average Phred quality (Q score) of at least 23, as well as perfect sequence matches to barcodes. For reads corresponding to each orthologue, the degenerate region was extracted. All extracted regions were then grouped and analysed with Weblogo[45].

Cell culture and transfection. Human embryonic kidney 293FT (Life Technologies), Neuro-2a (N2a), and Hepa1-6 (ATCC) cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS (HyClone), 2 mM GlutaMAX (Life Technologies), 100 U ml^{-1} penicillin, and 100 μg ml^{-1} streptomycin at 37 °C with 5% CO_2 incubation.

Cells were seeded into 24-well plates (Corning) one day before transfection at a density of 240,000 cells per well, and transfected at 70–80% confluency using Lipofectamine 2000 (Life Technologies) following the manufacturer's recommended protocol. For each well of a 24-well plate, a total of 500 ng DNA was used. For ChIP and BLESS, a total of 4.5 million cells are seeded the day before transfection into a 100-mm plate, and a total of 20 μg DNA was used.

DNA isolation from cells and tissue. Genomic DNA was extracted using the QuickExtract DNA Extraction Solution (Epicentre). Briefly, pelleted cells were resuspended in QuickExtract solution and incubated at 65 °C for 15 min, 68 °C for 15 min, and 98 °C for 10 min (ref. 8). Genomic liver DNA was extracted from bulk tissue fragments using a microtube bead mill homogenizer (Beadbug, Denville Scientific) by homogenizing approximately 30–50 mg of tissue in 600 μl of DPBS (Gibco). The homogenate was then centrifuged at 2,000 to 3,000*g* for 5 min at 4 °C and the pellet was resuspended in 300–600 μl QuickExtract DNA Extraction Solution (Epicentre) and incubated as above.

Indel analysis and guide:target base-pairing mismatch search. Indel analyses by SURVEYOR assay and targeted deep sequencing were carried out and analysed as previously described[8,22]. The methods for identification of potential off-target sites for SpCas9 based on Watson–Crick base-pairing mismatch between guide RNA and target DNA has been previously described[22], and adapted for SaCas9 by considering NNGRR for possible off-target PAMs. Alignment was manually adjusted to allow for insertion and deletion mismatches in the guide:target heteroduplex[29,30].

Chromatin immunoprecipitation and analysis. Cells were passaged at 24 h post-transfection into a 150-mm dish, and fixed for ChIP processing at 48 h post-transfection. For each condition, 10 million cells are used for ChIP input, following experimental protocols and analyses as previously described[31] with the following modifications: instead of pairwise peak-calling, ChIP peaks were only required to be enriched over both 'empty' controls (dSpCas9 only, dSaCas9 only) as well as the other Cas9/other sgRNA sample (for example, SpCas9/EMX-sg2 peaks must be enriched over SaCas9/EMX-sg1 peaks in addition to the empty controls). This was done to avoid filtering out of real peaks present in two related samples as much as possible.

To identify off-targets ranked by motif or sequence similarity to guide, motif scores for ChIP peaks were calculated as follows: for a given ChIP peak, the 100-nucleotide interval around the peak summit, the target sequence, and a given sgRNA guide region of length *L*, the query, an alignment score is calculated for every subsequence of *L* in the target. The subsequence with the highest score is reported as the best match to the query. For each subsequence alignment, the score calculation begins at the 5′ end of the query. For each position in the alignment, 1 is added or subtracted for match or mismatch between the query and target, respectively. If the score becomes negative, it is set to 0 and the calculation continued for the remainder of the alignment. The score at the 3′ end of the query is reported as the final score for the alignment. MACS scores $= -10\log(P$ value relative to the empty control) are determined as previously described[48]. For unbiased determination of PAM from ChIP peaks, the peaks were analysed for the best match by motif score to the guide region only within 50 nucleotides of the peak summit; the alignment was extended for 10 nucleotides at the 3′ end and visualized using Weblogo[45].

To calculate the motif score threshold at which false discovery rate < 0.1 for each sample, 100-nucleotide sequences centred around peak summits were shuffled while preserving dinucleotide frequency. The best match by motif score to the guide+PAM (NGG for SpCas9, NNGRRT for SaCas9) in these shuffled sequences was then found. The score threshold for false discovery rate < 0.1 was defined as the score such that less than 10% of shuffled peaks had a motif score above that score threshold.

BLESS for DSB detection. Cells are harvested at 24 h post-transfection, then processed as previously described[33] with the following alterations: a total of 10 million cells are fixed for nuclei isolation and permeabilization, and treated with Proteinase K for 4 min at 37 °C before inactivation with PMSF. All deproteinized nuclei are used for DSB labelling with 100 mM of annealed proximal linkers overnight. After Proteinase K digestion of labelled nuclei, chromatin was mechanically sheared with a 26G needle before sonication (BioRuptor, 20 min on high, 50% duty cycle). 20 μg of sheared chromatin are captured on streptavidin beads, washed, and ligated to 200 mM of distal linker. Linker hairpins are then cleaved off with I-SceI digestion for 1 h at 37 °C, and products PCR-enriched for 18 cycles before proceeding to library preparation with TruSeq Nano LT Kit (Illumina). For the negative control, cells mock transfected with Lipofectamine 2000 and pUC19 DNA were parallel processed through the assay.

BLESS analysis. Fastq files were demultiplexed, and 30-bp genomic sequences were separated from the BLESS ligation handles for alignment. Bowtie was used to map the genomic sequences to hg19 or mm9, allowing for a maximum of 2 mismatches. Following alignment, reads from all bio-replicates for an individual sample were first pooled, and then nearest neighbour clustering was performed with a 30-bp moving window to identify regions of enrichment across the genome. Within each cluster, the pairwise distance was calculated between all forward and reverse read strand mappings (Extended Data Fig. 7b, c). Pairwise distance distributions were used to filter out wide and poorly defined DSB clusters from the well-defined DSB clusters characteristically found at Cas9-induced cleavage sites (see Supplementary Information). Finally, we adjusted the count of predicted Cas9-induced DSBs at a given locus by using a binomial model to calculate the maximum-likelihood estimate of peak enrichment in the Cas9-sgRNA treated sgRNAs given BLESS measurements from an untreated negative control. After the maximum-likelihood estimate calculation, a list of loci ranked by their DSB scores could be obtained and plotted (Fig. 3b, Extended Data Fig. 8). Additional descriptions can be found in Supplementary Information.

The top-ranking ~30 sites from the list of Cas9 induced DSB clusters were sequenced for indel formation (Extended Data Fig. 8; validated targets in Fig. 3d). Within these loci, PAMs and regions of target homology were identified by first searching all PAM sites within a ± 50 bp window around the DSB cluster, then selecting the adjacent sequence with fewest mismatches to the target sequence.

Code availability. BLESS analysis code is available at https://github.com/fengzhanglab/BLESS.

Virus production and titration. For in-house viral production, 293FT cells (Life Technologies) were maintained as described above in 150 mm plates. For each transfection, 8 μg of pAAV8 serotype packaging plasmid, 10 μg of pDF6 helper plasmid, and 6 μg of AAV2 plasmid carrying the construct of interest were added to 1mL of serum-free DMEM. 125 μl of PEI "Max" solution (1 mg ml^{-1}, pH = 7.1) was then added to the mixture and incubated at room temperature for 5 to 10 s. After incubation, the mixture was added to 20 ml of warm maintenance media and applied to each dish to replace the old growth media. Cells were harvested between 48 h and 72 h post transfection by scraping and pelleting by centrifugation. The AAV2/8 (AAV2 inverted terminal repeat (ITR) vectors pseudo-typed with AAV8 capsid) viral particles were then purified from the pellet according to a previously published protocol[49].

High titre and purity viruses were also produced by vector core facilities at Children's Hospital Boston and Massachusetts Eye and Ear Infirmary (MEEI). These AAV vectors were then titred by real-time qPCR using a customized TaqMan probe against the transgene, and all viral preparations were titre-matched across different batches and production facilities before experiments. The purity of AAV vector was further verified by SDS–PAGE.

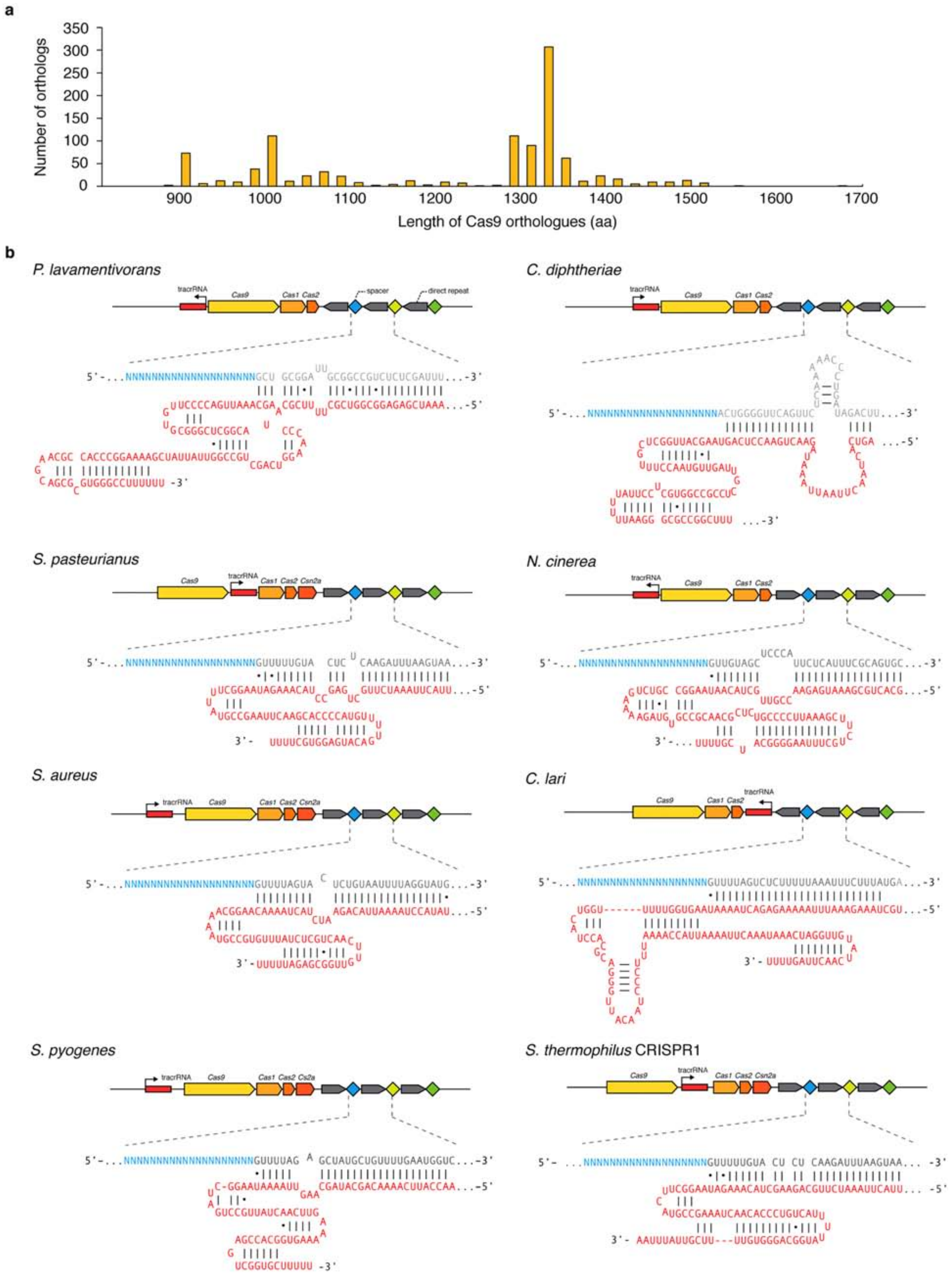
Animal injection and processing. All mice cohorts were maintained at animal facility with standard diet and housing following IRB-approved protocols. AAV vector was delivered to 5–6 week old male C57/BL6 mice intravenously via lateral tail vein injection. All dosages of AAV were adjusted to 100 μl or 200 μl with sterile phosphate buffered saline (PBS), pH 7.4 (Gibco) before the injection. Animals were not immunosuppressed or otherwise handled differently before injection or during the course of the experiment except the pre-bleed fasting as noted below. The animals were randomized to the different experimental conditions, with the investigator not blinded to the assignments.

To track the serum levels of Pcsk9 and total cholesterol, animals were fasted overnight for 12 h before blood collection by saphenous vein bleeds (no more than 100 μl or 10% of total blood volume per week). Multiple bleeds were made before tail vein delivery of AAV vector or control to collect pre-injection samples and to habituate the animals to handling during the procedure. After the blood was allowed to clot at room temperature, the serum was separated by centrifugation and stored at −20 °C for subsequent analysis. For terminal procedures to collect liver tissue and larger serum volumes for chemistry panels, mice were euthanized by carbon dioxide inhalation. Subsequently, blood was collected via cardiac puncture. Transcardial perfusion with 30 ml PBS removed the remaining blood, after which liver samples were collected. The median lobe of liver was removed and fixed in 10% neutral buffered formalin for histological analysis, while the remaining lobes were sliced in small blocks of size less than $1 \times 1 \times 3\ mm^3$ and frozen for subsequent DNA or protein extraction.

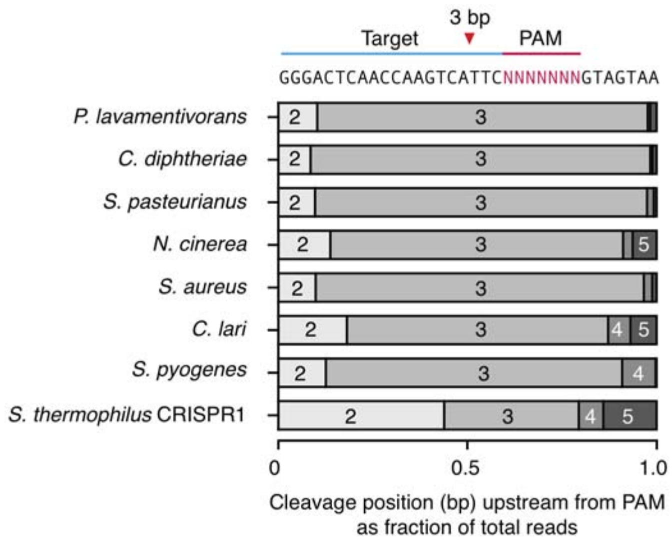
Histology and serum analysis. Following tissue harvesting as described above, flash-frozen mouse liver samples were embedded in OCT compound (Tissue Tek, Cat # 4583), snap-frozen, and stored at −80 °C before processing. Frozen tissues were cryosectioned at 4 μm in thickness and stained with Oil Red O following manufacturer's recommended protocol. Liver histology was assessed by H&E staining sections of 10% neutral buffer formalin fixed liver sections.

Serum levels of Pcsk9 were determined by ELISA using the Mouse Proprotein Convertase 9/PCSK9 Quantikine ELISA Kit (MPC-900, R&D Systems), following the manufacturer's instructions. Total cholesterol levels were measured using the Infinity Cholesterol Reagent (Thermo Fisher) per the manufacturer's instructions. Serum ALT, AST, albumin and total bilirubin were measured by an Olympus AU5400 (IDEXX Memphis, TN).

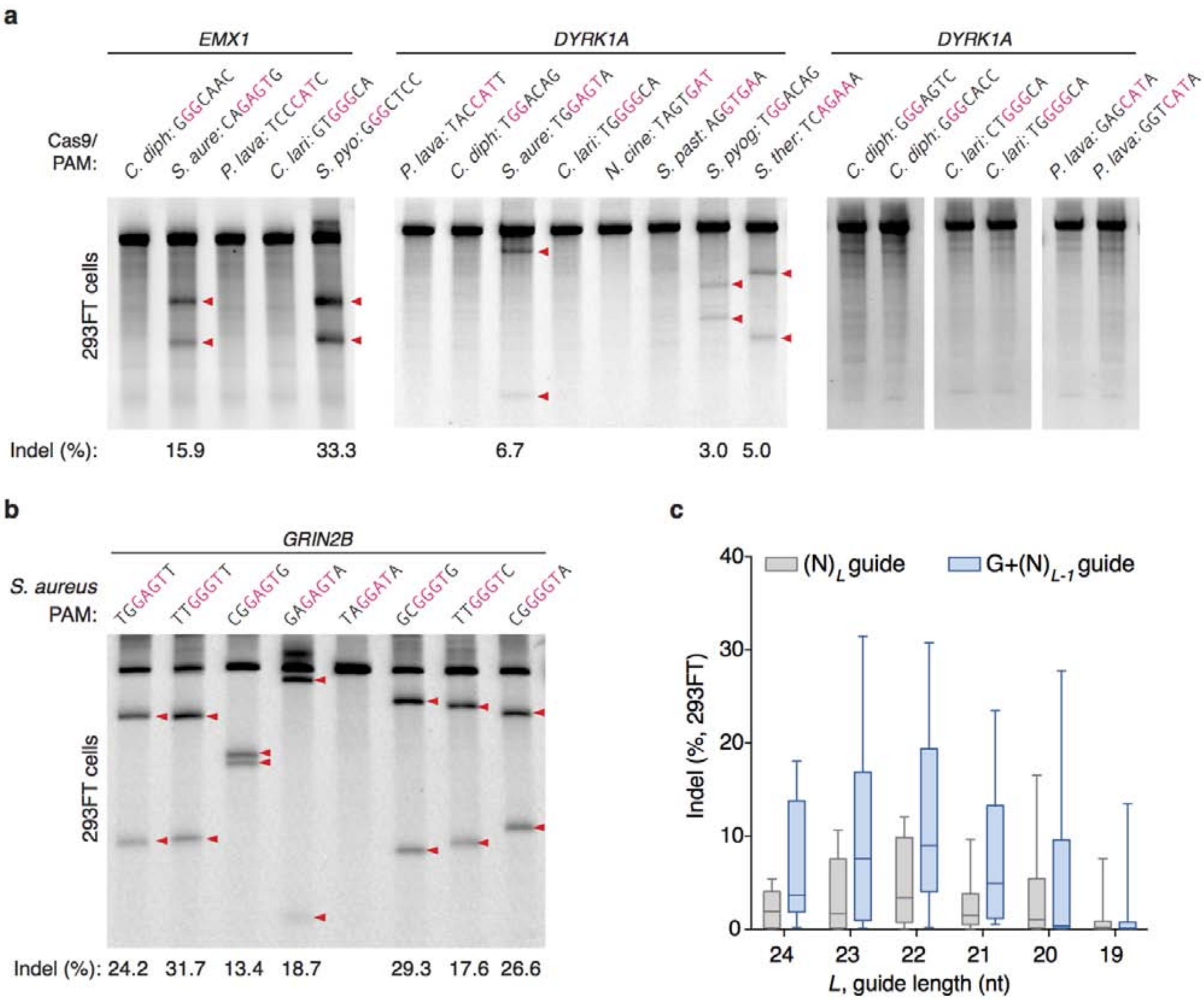
46. Gautheret, D. & Lambert, A. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J. Mol. Biol.* **313,** 1003–1011 (2001).
47. Macke, T. J. *et al.* RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res.* **29,** 4724–4735 (2001).
48. Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9,** R137 (2008).
49. Veldwijk, M. R. *et al.* Development and optimization of a real-time quantitative PCR-based method for the titration of AAV-2 vector stocks. *Mol. Ther.* **6,** 272–278 (2002).
50. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31,** 3406–3415 (2003).



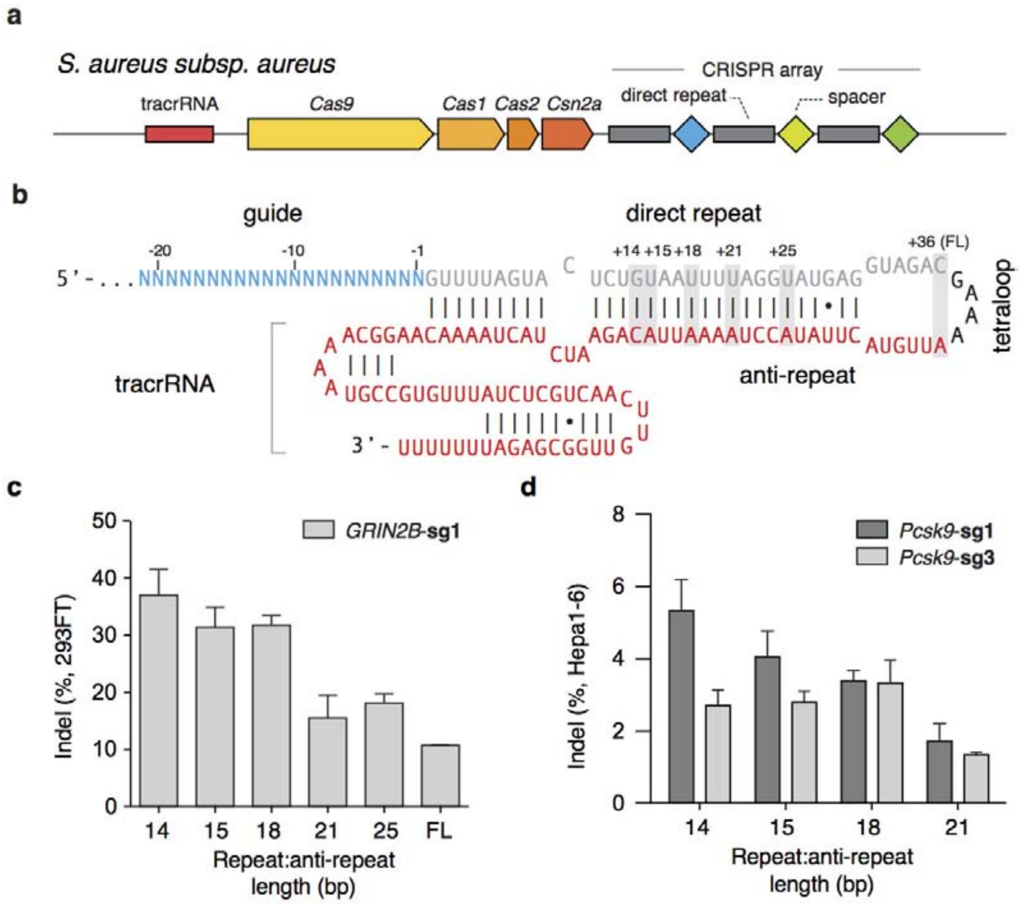
Extended Data Figure 1 | Selection of Type II CRISPR-Cas loci from eight bacterial species. a, Distribution of lengths for Cas9 > 600 Cas9 orthologues[19]. **b**, Schematic of Type II CRISPR-Cas loci and sgRNA from eight bacterial species. Spacer or 'guide' sequences are shown in blue, followed by direct repeats (grey). Predicted tracrRNAs are shown in red, and folded based on the Constraint Generation RNA folding model[50].



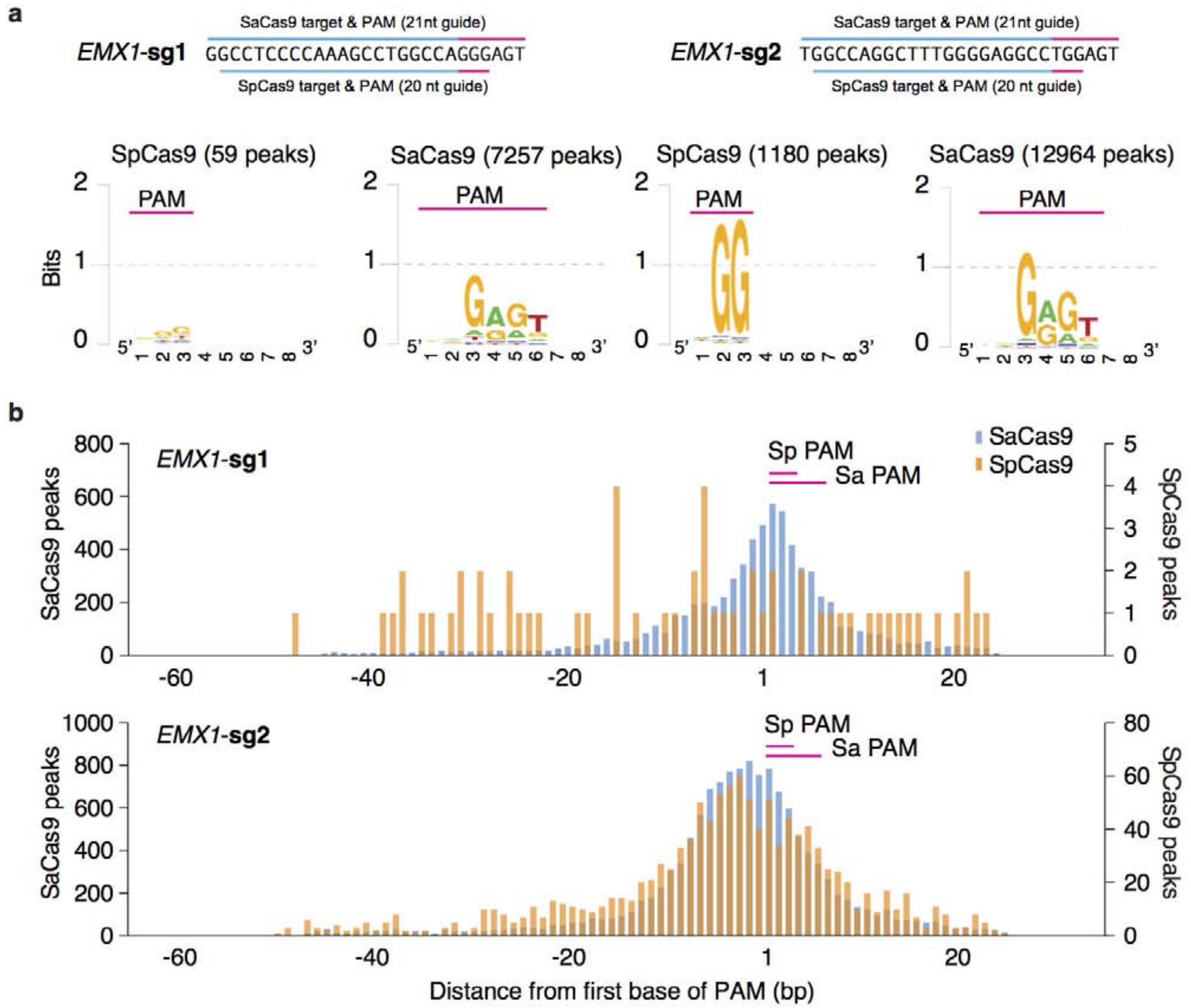
Extended Data Figure 2 | Cas9 orthologue cleavage pattern *in vitro*. Stacked bar graph indicates the fraction of targets cleaved at 2, 3, 4, or 5 bp upstream of PAM for each Cas9 orthologue; most Cas9 enzymes cleave stereotypically at 3 bp upstream of PAM (red triangle).



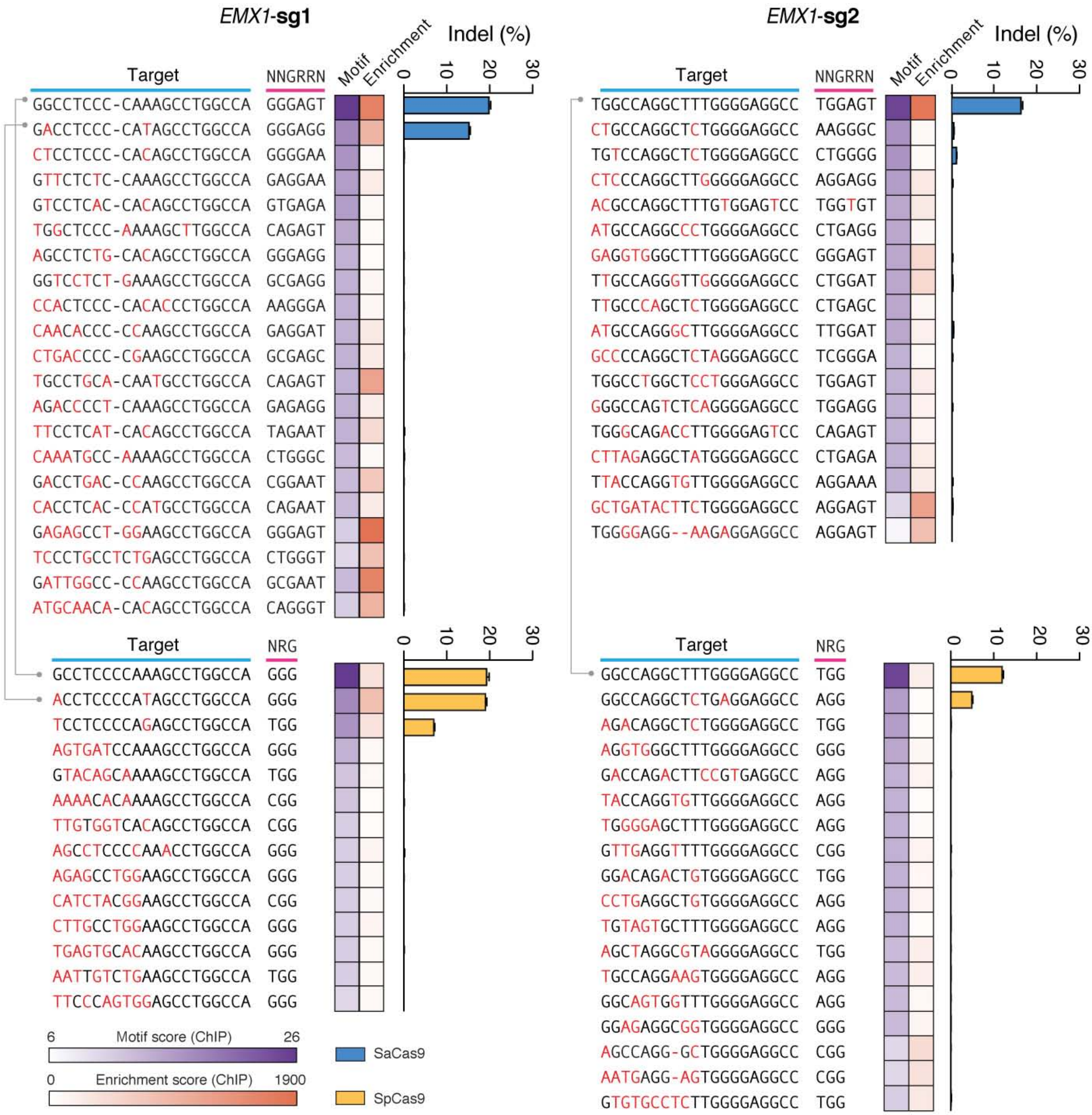
Extended Data Figure 3 | Test of Cas9 orthologue activity in 293FT cells. **a**, SURVEYOR assays showing indel formation at human endogenous loci from co-transfection of Cas9 orthologues and sgRNA. PAM sequences for individual targets are shown above each lane, with the consensus region for each PAM highlighted in red. Red triangles indicate cleaved fragments. **b**, SaCas9 generates indels efficiently for a multiple targets. **c**, Box-whisker plot of indel formation as a function of SaCas9 guide length L, with unaltered guides (perfect match of L nucleotides, grey bars) or replacement of the 5′-most base of guide with guanine ($G + L - 1$ nucleotides, blue bars) ($n = 8$ guides).



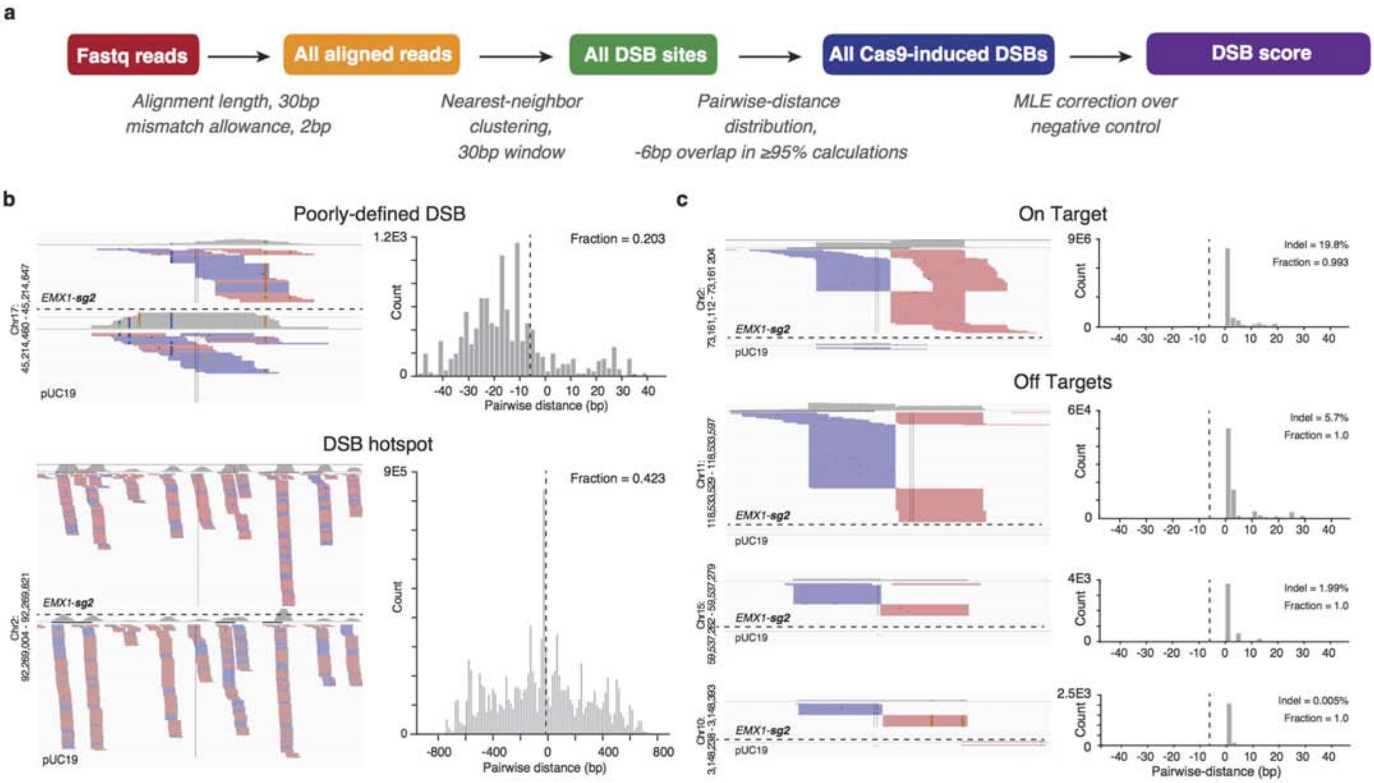
Extended Data Figure 4 | Optimization of SaCas9 sgRNA scaffold in mammalian cells. **a**, Schematic of the *Staphylococcus aureus subspecies aureus* CRISPR locus. **b**, Schematic of SaCas9 sgRNA with 21-nucleotide guide, crRNA repeat (grey), tetraloop (black) and tracrRNA (red). The number of crRNA repeat to tracrRNA anti-repeat base-pairing is indicated above the grey boxes. SaCas9 cleaves targets with varying repeat:anti-repeat lengths in **c**, HEK 293FT and **d**, Hepa1-6 cell lines. ($n = 3$, error bars show s.e.m.)



Extended Data Figure 5 | Genome-wide binding by Cas9-chromatin immunoprecipitation (dCas9-ChIP). **a**, Unbiased identification of PAM motif for dSaCas9 and dSpCas9. Peaks were analysed for the best match by motif score to the guide region only within 50 nucleotides of the peak summit. The alignment was extended for 10 nucleotides at the 3′ end and visualized using Weblogo. Numbers in parentheses indicate the number of called peaks. **b**, Histograms show the distribution of the peak summit relative to motif for dSaCas9 and dSpCas9. Position 1 on *x* axis indicates the first base of PAM.

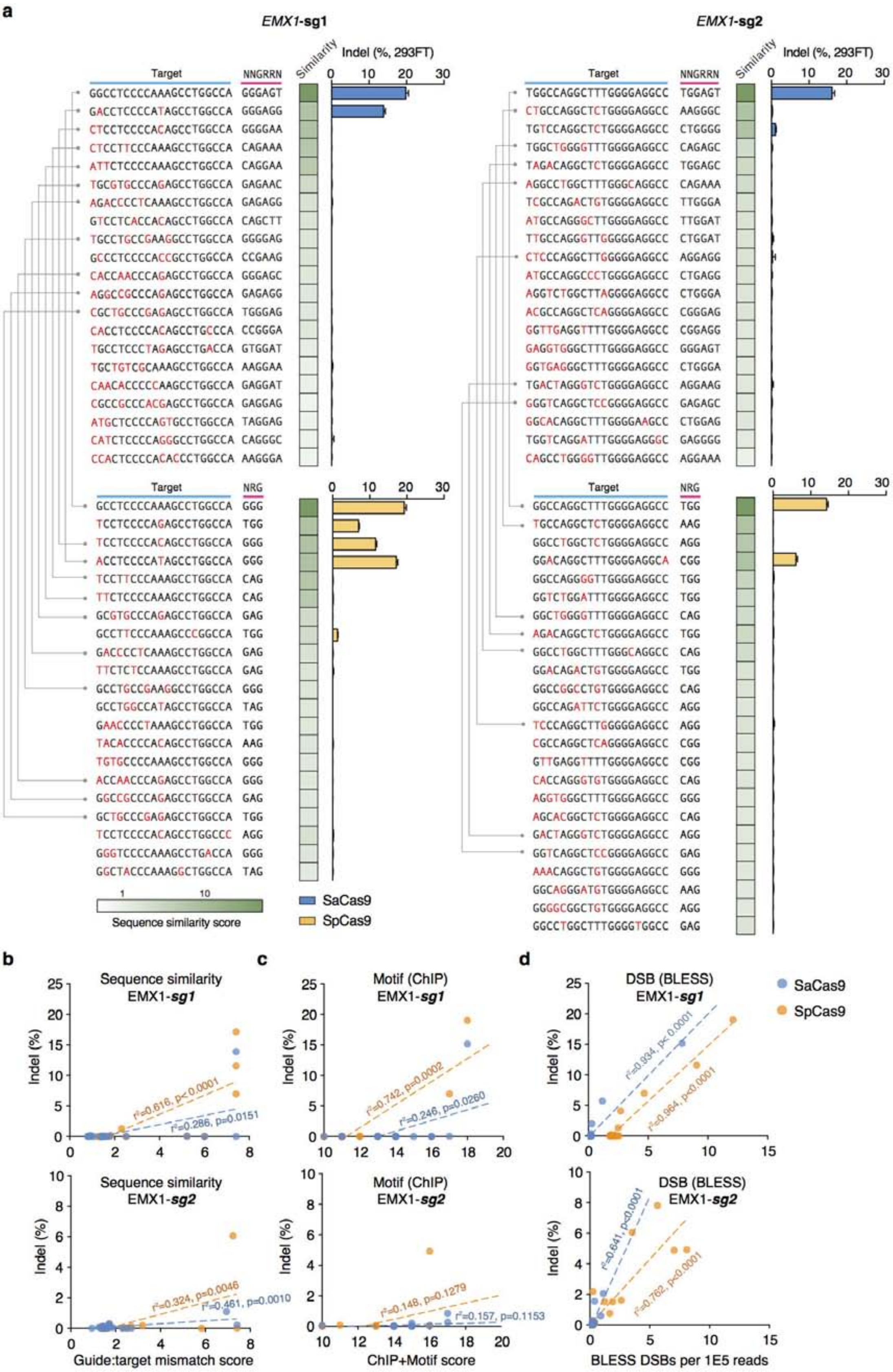


Extended Data Figure 6 | Indel measurements at candidate off-target sites based on ChIP. Indels at top off-target sites predicted by dCas9-ChIP for each Cas9 and sgRNA pair, based on ChIP peaks ranked by sequence similarity of the genomic loci to the guide motif (heat map in purple), or *P* value of ChIP enrichment over control (heat map in red). Lines connect the common targets (*EMX1*) and off-targets between the two Cas9 enzymes.

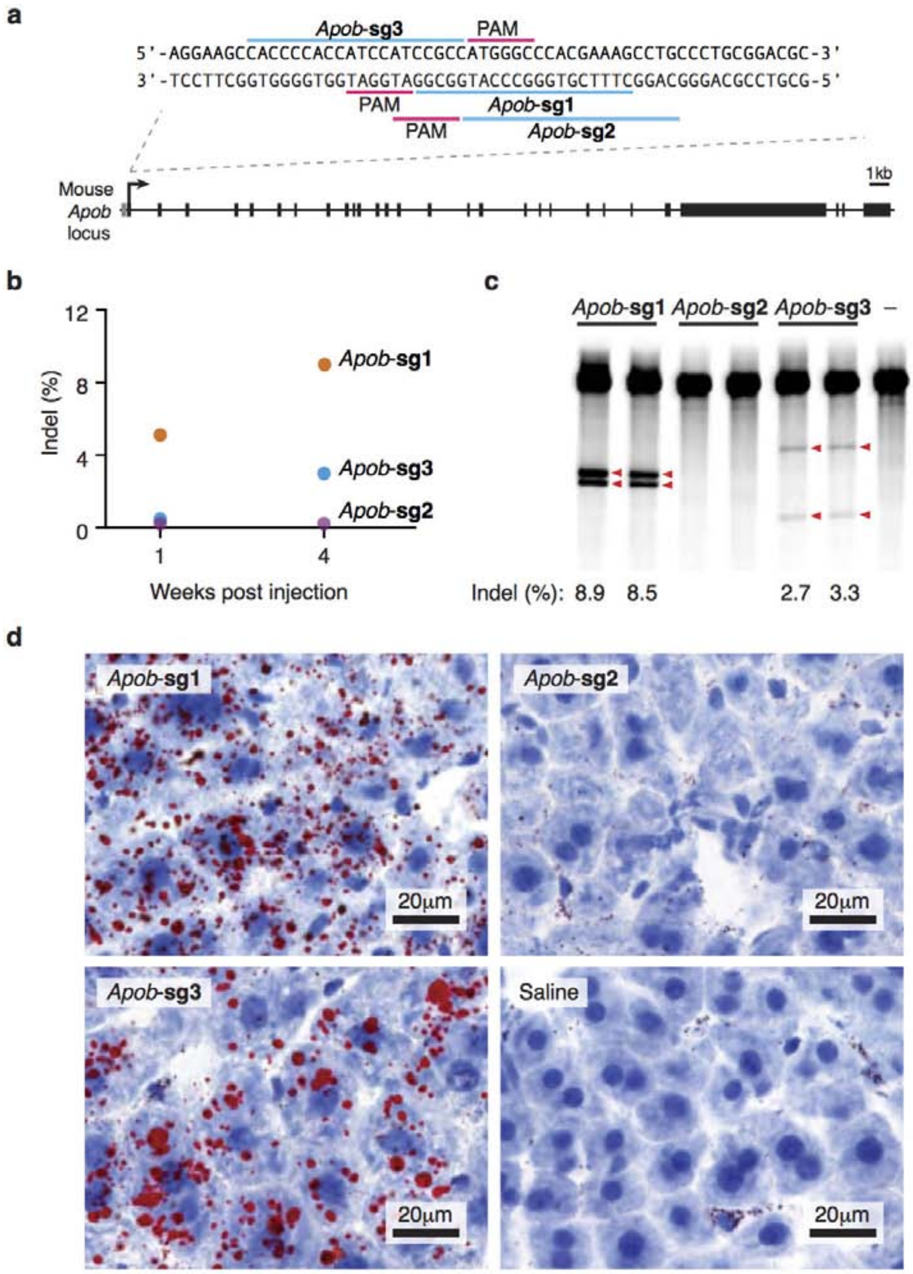


Extended Data Figure 7 | Analysis pipeline of sequencing data from BLESS. **a**, Overview of the data analysis pipeline starting from the raw sequencing reads. Representative sequencing read mappings and corresponding histograms of the pairwise distances between all the forward orientation (red) reads and reverse orientation (blue) reads, displayed for representative **b**, DSB hotspots and poorly defined DSB sites and **c**, Cas9-induced DSBs with detectable indels. Fraction of pairwise distances between reads overlapping by no more than 6 bp (dashed vertical line) are indicated over histogram plots.

Extended Data Figure 8 | Indel measurements at off-target sites based on DSB scores. List of top off-target sites ranked by DSB scores for each Cas9 and sgRNA pair. Indel levels are determined by targeted deep sequencing. Blue triangles indicate positions of peak BLESS signal, and where present, PAMs and targets with sequence homology to the guide are highlighted. Lines connect the common on-targets (*EMX1*) and off-targets between the two Cas9 enzymes. N.D., not determined.



Extended Data Figure 9 | Indel measurements of top candidate off-target sites based on sequence similarity score. Off-targets are predicted based on sequence similarity to on-target, accounting for number and position of Watson–Crick base-pairing mismatches as previously described[22]. NNGRR and NRG are used as potential PAMs for SaCas9 and SpCas9, respectively. Lines connect the common targets (*EMX1*) and off-targets between the two Cas9 enzymes. Correlation plots between indel percentages and **b**, prediction based on sequence similarity, **c**, ChIP peaks ranked by motif similarity, or **d**, DSB scores for top ranking off-target loci. Trendlines, r^2, and P values are calculated using ordinary least squares.



Extended Data Figure 10 | SaCas9 targeting *Apob* locus in the mouse liver. **a**, Schematics illustrating the mouse *Apob* gene locus and the positions of the three guides tested. **b**, Experimental time course and **c**, SURVEYOR assay showing indel formation at target loci after intravenous injection of AAV2/8 carrying thyroxine-binding globulin (TBG) promoter-driven SaCas9 and U6-driven guide at 2×10^{11} total genome copies ($n = 1$ animal each). **d**, Oil-red staining of liver tissue from AAV- or saline-injected animals. Male C56BL/6 mice were injected at 8 weeks of age and analysed 4 weeks post injection.

doi:10.1038/nature14362

Fatty acid carbon is essential for dNTP synthesis in endothelial cells

Sandra Schoors[1,2*], Ulrike Bruning[1,2*], Rindert Missiaen[1,2], Karla C. S. Queiroz[1,2], Gitte Borgers[1,2], Ilaria Elia[3,4], Annalisa Zecchin[1,2], Anna Rita Cantelmo[1,2], Stefan Christen[3,4], Jermaine Goveia[1,2], Ward Heggermont[5], Lucica Godde[1,2], Stefan Vinckier[1,2], Paul P. Van Veldhoven[6], Guy Eelen[1,2], Luc Schoonjans[1,2], Holger Gerhardt[7,8,9], Mieke Dewerchin[1,2], Myriam Baes[10], Katrien De Bock[1,2,11], Bart Ghesquière[1,2], Sophia Y. Lunt[12], Sarah-Maria Fendt[3,4] & Peter Carmeliet[1,2]

The metabolism of endothelial cells during vessel sprouting remains poorly studied. Here we report that endothelial loss of CPT1A, a rate-limiting enzyme of fatty acid oxidation (FAO), causes vascular sprouting defects due to impaired proliferation, not migration, of human and murine endothelial cells. Reduction of FAO in endothelial cells did not cause energy depletion or disturb redox homeostasis, but impaired *de novo* nucleotide synthesis for DNA replication. Isotope labelling studies in control endothelial cells showed that fatty acid carbons substantially replenished the Krebs cycle, and were incorporated into aspartate (a nucleotide precursor), uridine monophosphate (a precursor of pyrimidine nucleoside triphosphates) and DNA. CPT1A silencing reduced these processes and depleted endothelial cell stores of aspartate and deoxyribonucleoside triphosphates. Acetate (metabolized to acetyl-CoA, thereby substituting for the depleted FAO-derived acetyl-CoA) or a nucleoside mix rescued the phenotype of CPT1A-silenced endothelial cells. Finally, CPT1 blockade inhibited pathological ocular angiogenesis in mice, suggesting a novel strategy for blocking angiogenesis.

Angiogenesis relies on the proliferation and migration of endothelial cells[1]. The importance of endothelial cell proliferation for expansive growth of the vascular network has long been recognized[2]. While numerous molecules regulating vessel sprouting have been identified[3], little is known about the role of metabolism. We recently reported that endothelial cells generate 85% of their ATP for vessel sprouting via glycolysis[4]. Fatty acid oxidation (FAO) has been linked in various cell types to ATP production and to ROS scavenging during cellular stress, but apart from a few earlier *in vitro* reports[5], the role and importance of FAO in endothelial cells during angiogenesis *in vivo* is undefined.

By shuttling long-chain fatty acids into mitochondria, carnitine palmitoyltransferase 1 (CPT1) constitutes a rate-limiting step of FAO. Oxidation of palmitate generates acetyl-CoA, which fuels the TCA cycle. Apart from generating ATP, the TCA cycle also provides precursors for macromolecule synthesis, necessary for proliferation. However, fatty acids have not yet been shown to function as carbon sources for biosynthetic processes. Here we elucidated the role of FAO in endothelial cells during angiogenesis, and studied how FAO determines endothelial cell behaviour.

FAO stimulates vessel sprouting via EC proliferation

To study the role of mitochondrial FAO in vessel sprouting, we silenced CPT1A, the most abundant CPT1 isoform in human umbilical venous endothelial cells (HUVECs), which lowered levels of *CPT1A* mRNA and protein and reduced FAO flux (Extended Data Fig. 1a–f). In contrast, silencing of CPT1C, expressed at lower levels (Extended Data Fig. 1a), did not affect FAO (Extended Data Fig. 1g). As similar data were obtained in other endothelial cell subtypes (Extended Data Fig. 1d, h), we used HUVECs (hereafter referred to as endothelial cells (ECs)) for our study.

In EC spheroids, CPT1A silencing (CPT1A knockdown ($CPT1A^{KD}$)) decreased vessel sprout length and numbers (Fig. 1a–c and Extended Data Fig. 1i). This defect was due to decreased EC proliferation, as $CPT1A^{KD}$ reduced proliferation and had only minimally additive effects on mitomycin-C-treated mitotically inactivated ECs (Fig. 1c–f and Extended Data Fig. 1i, j). By contrast, $CPT1A^{KD}$ did not affect EC migration or motility (Fig. 1g–i and Extended Data Fig. 1k). Similar results were obtained when silencing long-chain acyl-CoA dehydrogenase (*ACADVL*), another FAO gene (Extended Data Fig. 1l–o). Additional evidence for a role of FAO in vessel sprouting was provided by overexpression of CPT1A ($CPT1A^{OE}$), which yielded opposite results to those obtained by $CPT1A^{KD}$ (Extended Data Fig. 1p–t). Thus, CPT1A-driven FAO regulates EC proliferation during vessel sprouting.

To study the effects of endothelial CPT1A deficiency on vessel formation *in vivo*, we used $Cpt1a^{lox/lox}$ mice intercrossed with *VE-cadherin(PAC)-Cre*^{ERT2} mice ($CPT1A^{\Delta EC}$ mice) (Extended Data Fig. 1u). Isolectin-B4 staining of retinal vessels of transgenic pups at postnatal day (P) 5 revealed that EC loss of CPT1A diminished the number of vascular branch points, and reduced radial expansion of the vascular plexus (Fig. 2a–e). This was not caused by increased vessel regression (Fig. 2f–h) but by reduced EC proliferation (Fig. 2i–k). Furthermore, $CPT1A^{\Delta EC}$ mice had normal numbers of filopodia (Fig. 2l–n) and vessel maturation (Fig. 2o, p and Extended Data Fig. 1v). Thus, impaired angiogenesis in $CPT1A^{\Delta EC}$ mice was due to EC proliferation defects.

[1]Laboratory of Angiogenesis and Neurovascular link, Department of Oncology, KU Leuven, B-3000 Leuven, Belgium. [2]Laboratory of Angiogenesis and Neurovascular Link, Vesalius Research Center, VIB, B-3000 Leuven, Belgium. [3]Laboratory of Cellular Metabolism and Metabolic Regulation, Department of Oncology, KU Leuven, B-3000 Leuven, Belgium. [4]Laboratory of Cellular Metabolism and Metabolic Regulation, Vesalius Research Center, VIB, B-3000 Leuven, Belgium. [5]Center for Molecular & Vascular Biology, Department of Cardiovascular Research, KU Leuven; Division of Clinical Cardiology, UZ Leuven, B-3000 Leuven, Belgium. [6]Laboratory of Lipid Biochemistry and Protein Interactions, KU Leuven, B-3000 Leuven, Belgium. [7]Vascular Patterning Laboratory, Department of Oncology, KU Leuven, B-3000 Leuven, Belgium. [8]Vascular Patterning Laboratory, Vesalius Research Center, VIB, B-3000 Leuven, Belgium. [9]Integrative Vascular Biology Laboratory, Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany. [10]Laboratory of Cell Metabolism, Department of Pharmaceutical and Pharmacological Sciences, KU Leuven, B-3000 Leuven, Belgium. [11]Exercise Physiology Research Group, Department of Kinesiology, KU Leuven, B-3001 Leuven, Belgium. [12]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA.
*These authors contributed equally to this work.

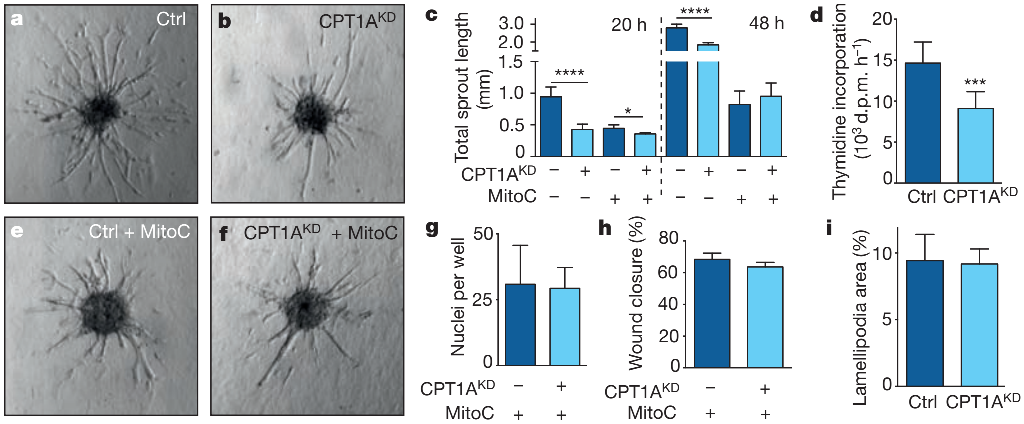


Figure 1 | FAO stimulates vessel sprouting via EC proliferation. **a**, **b**, Representative images (original magnification, 10×) of control (Ctrl) and CPT1A^{KD} EC spheroids. **c**, Total sprout length in control and CPT1A^{KD} EC spheroids treated with mitomycin C (MitoC) where indicated ($n = 3$). **d**, [^{3}H] thymidine incorporation in DNA in control and CPT1A^{KD} ECs ($n = 5$). d.p.m., disintegrations per minute. **e**, **f**, Representative images (original magnification, 10×) of MitoC-treated control and CPT1A^{KD} EC spheroids. **g**, Number of MitoC-treated control and CPT1A^{KD} ECs that traversed a Boyden chamber ($n = 4$; P = NS). **h**, Scratch wound assay using MitoC-treated control and CPT1A^{KD} ECs ($n = 4$; P = NS). **i**, Lamellipodia area in control and CPT1A^{KD} ECs ($n = 4$; P = NS). Data are mean ± s.e.m. of n independent experiments. Statistical test: mixed models (**c**, **d**, **g**–**i**). NS, not significant. *$P < 0.05$, ***$P < 0.001$, ****$P < 0.0001$.

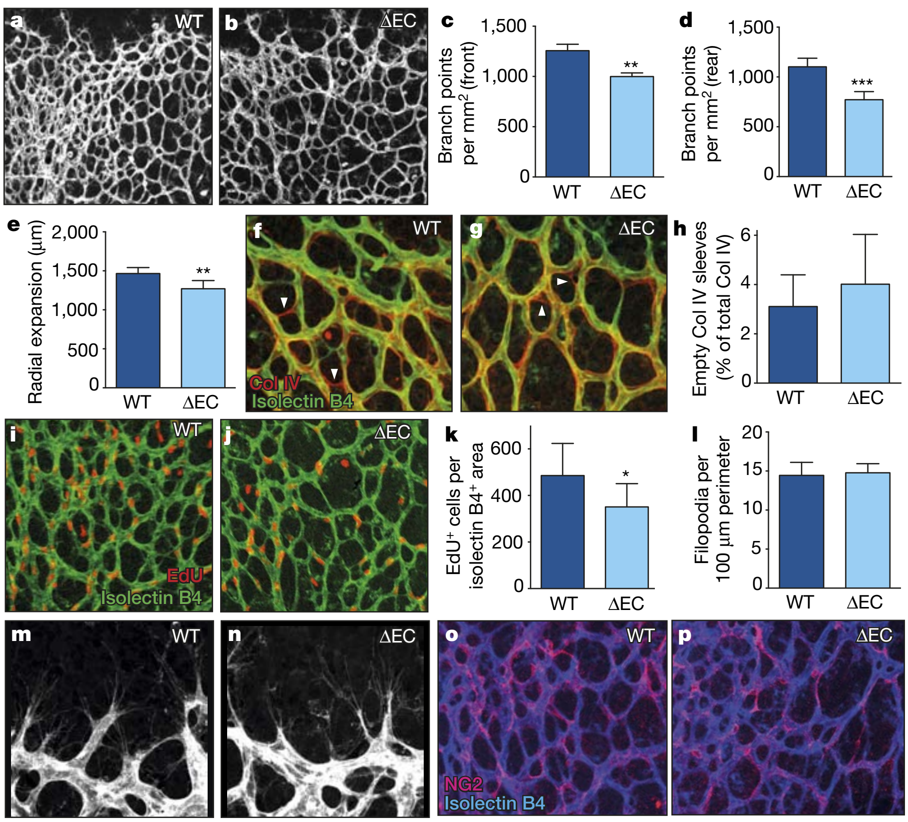


Figure 2 | CPT1A gene deletion in ECs causes vascular defects *in vivo*. **a**, **b**, Representative images (original magnification, ×20) of retinal vessels of wild-type (WT; **a**) and CPT1A$^{\Delta EC}$ (**b**) mice. **c**, **d**, Branch point quantification in WT and CPT1A$^{\Delta EC}$ mice in the front (**c**) and rear (**d**) of the retinal vasculature (5 litters, $n = 11$ pups for WT and CPT1A$^{\Delta EC}$). **e**, Retinal vascular outgrowth in wild-type and CPT1A$^{\Delta EC}$ mice (6 litters, $n = 13$ pups for WT and 18 for CPT1A$^{\Delta EC}$). **f**, **g**, Representative images (original magnification, 20×) of the retina stained for isolectin-B4 (green) and collagen IV (red) in WT (**f**) and CPT1A$^{\Delta EC}$ mice (**g**). **h**, Quantification of isolectin-B4$^-$ collagen IV$^+$ empty sleeves in WT and CPT1A$^{\Delta EC}$ pups (4 litters, $n = 8$ pups for WT and 14 for CPT1A$^{\Delta EC}$; P = NS). **i**, **j**, Representative images (original magnification, 10×) of retina stained for EdU (red) and isolectin-B4 (green) in WT (**i**) and CPT1A$^{\Delta EC}$ (**j**) mice. **k**, Quantification of EdU$^+$ ECs in WT and CPT1A$^{\Delta EC}$ mice (3 litters, $n = 9$ pups for WT and 6 for CPT1A$^{\Delta EC}$). **l**, Quantification of filopodia in WT and CPT1A$^{\Delta EC}$ mice (6 litters, $n = 20$ pups for WT and 16 for CPT1A$^{\Delta EC}$; P = NS). **m**, **n**, Representative images of filopodia (original magnification, 63×) in WT (**m**) and CPT1A$^{\Delta EC}$ (**n**) mice. **o**, **p**, Representative images (original magnification, 20×) of the retinal vasculature of WT (**o**) and CPT1A$^{\Delta EC}$ mice (**p**) stained for isolectin-B4 (blue) and the pericyte marker NG2 (pink). Data are mean ± s.e.m. of n individual mice. Statistical test: mixed models (**c**–**e**, **h**, **k**, **l**). NS, not significant. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

CPT1A^{KD} does not cause ATP or redox imbalance

To explore by which mechanism FAO regulates EC proliferation, we studied whether CPT1A^{KD} impaired EC proliferation by causing energy stress due to ATP depletion. However, CPT1A^{KD} did not lower ATP levels, nor did it affect the energy charge or ADP/ATP ratio (Fig. 3a, b and Extended Data Fig. 2a). This was not surprising, as FAO generated only 5% of the total amount of ATP in ECs[4]. Oxygen consumption (OCR) coupled to ATP synthesis was modestly reduced (Fig. 3c), but could not explain the sprouting defect, as blocking mitochondrial ATP production with oligomycin failed to affect sprouting of control and CPT1A^{KD} ECs, suggesting that CPT1A^{KD} impaired sprouting independently of mitochondrial ATP production (Fig. 3d and Extended Data Fig. 2b). Additional experiments confirmed that CPT1A^{KD} ECs were not in energy stress (Extended Data Fig. 2c, d).

We also explored whether CPT1A^{KD} impaired vessel sprouting by altering redox balance. CPT1A^{KD} increased ROS levels by only 20%, an increase reported to enhance EC proliferation[6], without affecting the percentage of oxidized glutathione or disturbing redox homeostasis (Fig. 3e, f). Furthermore, lowering ROS levels by using *N*-acetyl-cysteine (NAC) did not restore vessel sprouting upon CPT1A silencing (Fig. 3g and Extended Data Fig. 2e). Finally, CPT1A^{KD} did not compromise EC survival and did not increase levels of oxidative DNA damage markers (Extended Data Fig. 2f–j). Thus, CPT1A^{KD} did not impair vessel sprouting by inducing toxic ROS levels.

FAO is used for *de novo* synthesis of nucleotides

We thus considered a novel role for FAO in EC proliferation and explored whether FAO regulated the production of biomass building blocks. Supplementing EC monolayers with [U-^{13}C]palmitate or an algal [U-^{13}C]fatty acid mix revealed that carbons from fatty acids provided a substantial fraction of the total carbon fuelling the TCA cycle intermediates and TCA-cycle-derived amino acids; this fraction was similar to the contribution of carbons from glutamine and glycolytic carbon (Fig. 4a–c)[7]. This was unexpected as many cancer cell types rely almost exclusively on glucose and glutamine to fuel the TCA cycle[8]. CPT1A^{KD} also lowered the cellular pool size of citrate, aspartate and glutamate (Fig. 4d).

Since TCA intermediates are used for the synthesis of biomass precursors, and inhibition of FAO limited the supply of these TCA intermediates, we explored whether CPT1A^{KD} impaired protein and/or nucleotide synthesis. CPT1A^{KD} did not, however, impair *de novo* protein synthesis (Fig. 4e) and did not consistently alter intracellular amino acid levels (Fig. 4d). Also, CPT1A^{KD} still lowered EC proliferation when protein synthesis was reduced by cycloheximide (Fig. 4f), suggesting that a decrease of *de novo* protein synthesis did not prevent CPT1A silencing to establish its proliferation defect.

Given that aspartate is a precursor of nucleotides and its levels were reduced upon CPT1A^{KD}, we explored whether fatty-acid-derived carbons were used for *de novo* ribonucleotide synthesis. Label from [U-^{14}C]palmitate was incorporated into RNA in control ECs, and this process was decreased upon CPT1A^{KD} (Fig. 4g). However, despite this, CPT1A^{KD} ECs did not have reduced RNA levels and pools of ribonucleotides (rNTPs) (Fig. 4h, i), presumably because of compensatory production by salvage pathways. Thus, insufficient RNA synthesis did not cause the proliferation defect of CPT1A^{KD} ECs.

We therefore assessed whether impaired *de novo* deoxyribonucleotide synthesis caused the proliferation defect of CPT1A^{KD} ECs. Label from [U-^{14}C]palmitate was incorporated into DNA, and this was reduced upon CPT1A silencing (Fig. 4j). As CPT1A^{KD} also decreased *de novo* DNA synthesis (Fig. 1d), we hypothesized that FAO blockade reduced aspartate levels and thereby compromised *de novo* deoxyribonucleotide biosynthesis. Indeed, labelled carbons from [U-^{13}C]palmitate were incorporated in the pyrimidine precursors uridine monophosphate (UMP) and uridine-5′-triphosphate (UTP), and this incorporation was reduced upon CPT1A silencing (Fig. 4k, l and Extended Data Fig. 3a). Moreover, quantification of deoxyribonucleotide (dNTP) levels confirmed that CPT1A^{KD} lowered the intracellular levels of all pyrimidine (dCTP, dTTP) and purine deoxyribonucleotides (dATP, dGTP) needed for DNA replication (Fig. 4m). Thus, in contrast to the retained rNTP levels, CPT1A^{KD} depleted the pool of dNTPs.

To functionally confirm that the effect of CPT1A^{KD} depends on reduced *de novo* nucleotide synthesis, we inhibited *de novo* synthesis of nucleotides by methotrexate (MTX) and 5-fluorouracil (5FU), which reduced vessel sprouting in control but did not further reduce vessel sprouting in CPT1A^{KD} cells (Fig. 4n, o and Extended Data Fig. 3b, c). Thus, CPT1A^{KD} reduced sprouting mainly via inhibition of *de novo* deoxyribonucleotide synthesis.

To confirm the role of fatty-acid-derived carbons for nucleotide synthesis, we supplemented CPT1A^{KD} cells with acetate (which is converted to acetyl-CoA that can fuel the TCA cycle). Treatment of CPT1A^{KD} ECs

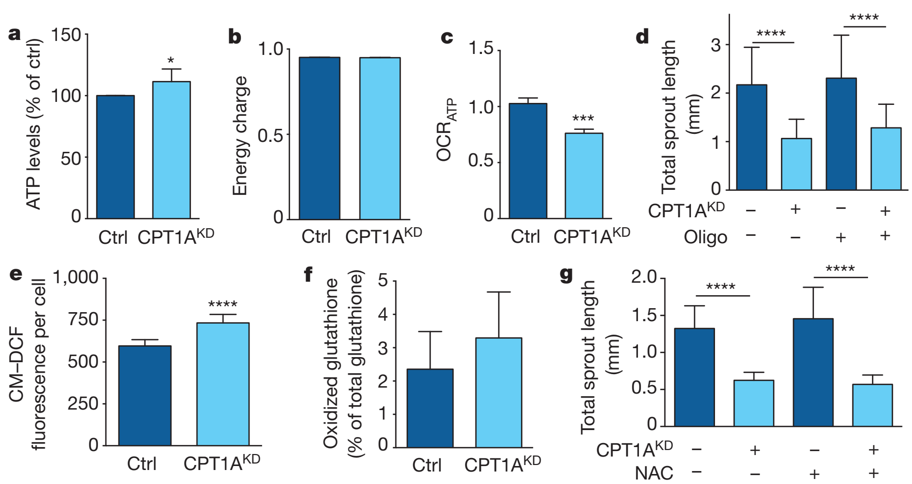


Figure 3 | CPT1A silencing does not cause ATP depletion or redox imbalance. **a**, Intracellular ATP levels in control and CPT1A^{KD} ECs ($n = 4$). **b**, Energy charge measurement (ratio) in control and CPT1A^{KD} ECs ($n = 3$; P = NS). **c**, ATP coupled oxygen consumption rate (OCR_{ATP}, expressed relative to control) in control and CPT1A^{KD} ECs ($n = 3$). **d**, Total sprout length upon oligomycin (oligo) treatment in control and CPT1A^{KD} EC spheroids ($n = 3$). **e**, Intracellular ROS levels in control and CPT1A^{KD} ECs as measured using the fluorescent ROS detection reagent chloromethyl-2′,7′-dichlorofluorescin diacetate, acetyl ester (CM-DCF) ($n = 6$). **f**, Oxidized glutathione levels as per cent of total glutathione, in control and CPT1A^{KD} ECs ($n = 3$; P = NS). **g**, Total sprout length upon *N*-acetyl-cysteine (NAC) treatment in control and CPT1A^{KD} EC spheroids ($n = 3$). Data are mean ± s.e.m. of n independent experiments. Statistical test: mixed models (**a**–**g**). NS, not significant. *$P < 0.05$, ***$P < 0.001$, ****$P < 0.0001$.

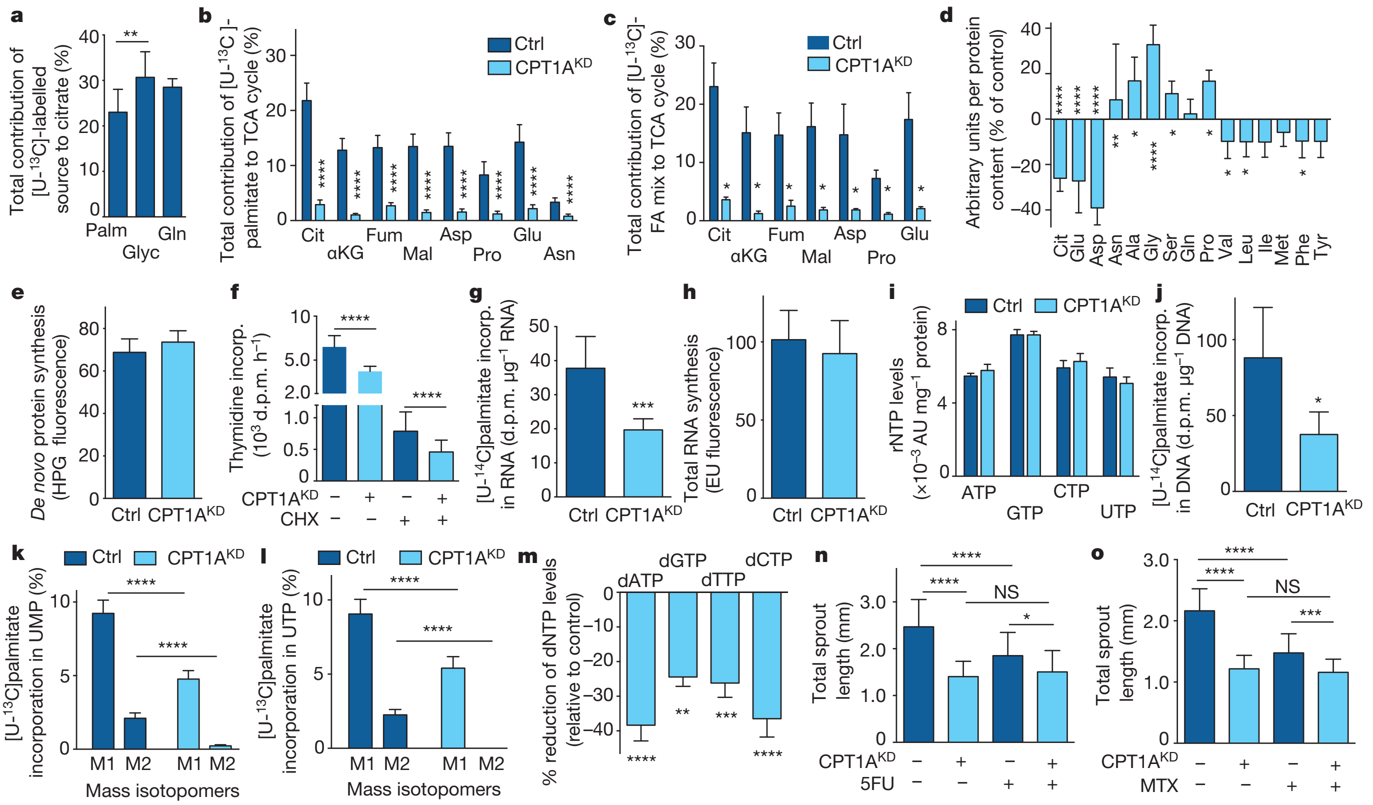


Figure 4 | CPT1A silencing reduces TCA replenishment and FAO is used for nucleotide synthesis. **a**, Total contribution of palmitate (Palm), glycolytic carbon (Glyc) and glutamine (Gln) to citrate ($n = 3$). **b**, **c**, Total contribution of [U-^{13}C]palmitate (**b**) or [U-^{13}C]algal fatty acid (FA) mix (**c**) to citrate (Cit), α-ketoglutarate (αKG), fumarate (Fum), malate (Mal), aspartate (Asp), proline (Pro), glutamate (Glu) or asparagine (Asn) in control and CPT1A^{KD} ECs ($n = 7$ for TCA intermediates, $n = 3$ for Asn and $n = 5$ for all other amino acids in **b**; $n = 3$ in **c**). **d**, Intracellular content of citrate (Cit), glutamate (Glu), aspartate (Asp), glutamine (Gln), isoleucine (Ile), methionine (Met), tyrosine (Tyr), asparagine (Asn), alanine (Ala), glycine (Gly), serine (Ser), proline (Pro), valine (Val), leucine (Leu) and phenylalanine (Phe) in control and CPT1A^{KD} ECs ($n = 8$ for Cit and Asp, $n = 6$ for Glut, $n = 4$ for Asn and Ser, $n = 5$ for Ala, Gly, Gln, Pro, Val, Leu, Ile, Met, Phe and Tyr). **e**, *De novo* protein synthesis in control and CPT1A^{KD} ECs ($n = 3$; $P =$ NS). HPG, L-homopropargylglycine. **f**, [^{3}H]thymidine incorporation upon cycloheximide (CHX) treatment in control and CPT1A^{KD} ECs ($n = 3$). **g**, [U-^{14}C]palmitate incorporation into RNA in control and CPT1A^{KD} ECs ($n = 5$). **h**, *De novo* RNA synthesis in control and CPT1A^{KD} ECs ($n = 3$; $P =$ NS). EU fluorescence (5′-ethynyl uridine) in arbitrary units. **i**, Intracellular levels of the rNTPs in control and CPT1A^{KD} ECs ($n = 4$; $P =$ NS). AU, arbitrary units. **j**, [U-^{14}C]palmitate incorporation into DNA in control and CPT1A^{KD} ECs ($n = 5$). **k**, **l**, Percentage M+1 and M+2 labelling from [U-^{13}C]palmitate in UMP (**k**) and UTP (**l**) in control and CPT1A^{KD} ECs ($n = 5$). M denotes mass of the unlabelled metabolite. **m**, Reduction of intracellular dNTP levels in CPT1A^{KD} versus control ECs ($n = 5$ for dATP and dCTP, $n = 4$ for dTTP and dGTP). **n**, **o**, Total sprout length upon 5-fluorouracil (5FU) (**n**) or methotrexate (MTX) (**o**) treatment in control and CPT1A^{KD} ECs ($n = 4$). Data are mean ± s.e.m. of n independent experiments. Statistical test: mixed models (**a**–**o**). NS, not significant. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

with acetate completely restored the levels of aspartate and dNTPs (Fig. 5a, b). Furthermore, CPT1A^{KD} and control cells incorporated comparable amounts of [U-^{13}C]acetate into UMP and UTP (Fig. 5c, d). Functionally, acetate rescued the sprouting defect of CPT1A^{KD} ECs (Fig. 5e and Extended Data Fig. 3d–g). Also, blockade of the ATP synthase did not prevent this rescue, indicating that acetate was not used for the production of ATP but instead for the synthesis of biomass (Extended Data Fig. 3h, i).

Supplementation of a mix of nucleosides or dNTPs also fully rescued the sprouting defect of CPT1A^{KD} vessels, indicating that FAO blockade impaired nucleotide synthesis (Fig. 5f and Extended Data Fig. 3j, k). Notably, acetate or the nucleoside mix were unable to rescue the CPT1A^{KD} sprouting defect upon mitotic inactivation of CPT1A^{KD} ECs, indicating that their effect relied on stimulating EC proliferation (Extended Data Fig. 3l; not shown). Thus, CPT1A-driven FAO is critical for *de novo* deoxyribonucleotide synthesis to ensure EC proliferation during sprouting.

Glucose partly compensates for FAO loss

Nucleotides can also be synthesized from glucose and glutamine[9], so we asked whether ECs would compensate for the loss of FAO using these nutrients. CPT1A^{KD} increased the oxidation of glucose but not of glutamine (Extended Data Fig. 3m, n), and increased glycolytic carbon incorporation into citrate and aspartate, whereas glutamine incorporation into citrate and aspartate was not (citrate), or only negligibly (aspartate), altered (Fig. 5g and Extended Data Fig. 3o). The relative activity of pyruvate dehydrogenase and pyruvate carboxylase was increased in CPT1A^{KD} ECs (Fig. 5h–j), which can contribute to the increase in glucose oxidation. Nonetheless, this compensation was insufficient to rescue the proliferation defect of CPT1A^{KD} ECs, indicating that fatty acids are irreplaceable for DNA replication in ECs.

Most cells do not use fatty acids for DNA synthesis

To assess the broader relevance of the role of FAO in deoxyribonucleotide synthesis, we studied [U-^{13}C]palmitate contribution to citrate in a panel of normal and malignant proliferating cells. Fibroblasts (and to a certain extent pericytes) but no other primary cells incorporated substantial amounts of fatty-acid-derived carbon into citrate (Fig. 6a). Except for breast cancer (T47D, MCF7) and lung cancer (A549) cell line (known to rely on FAO for ATP and NADPH production[10]), fatty acids contributed minimally to citrate production (Fig. 6a). Notably, even when these cancer cells incorporated fatty-acid-derived carbons into citrate, they incorporated them to a much lesser extent into DNA (Fig. 6b), implying that fatty-acid-derived carbons entering the TCA

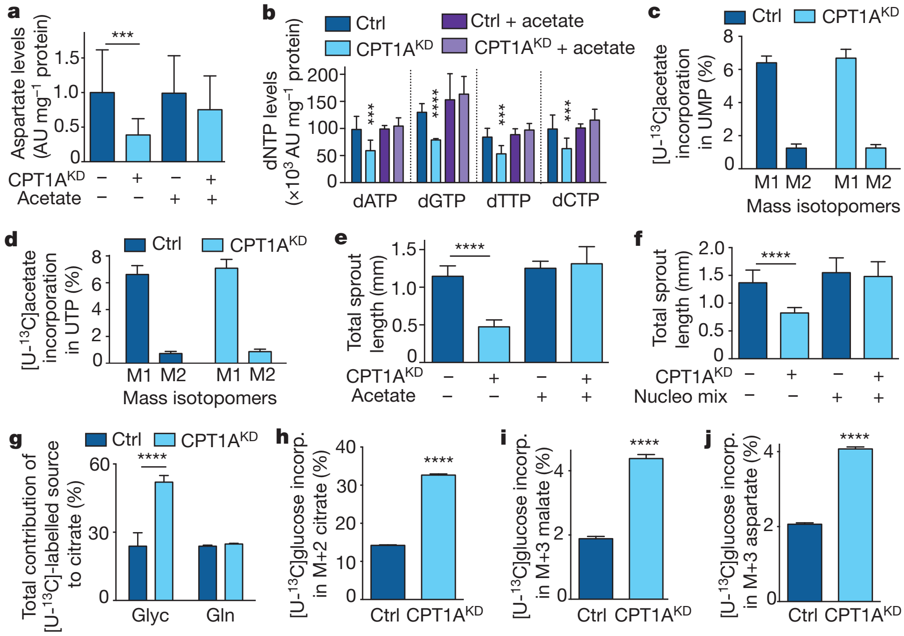


Figure 5 | Acetate or nucleosides rescue the CPT1A^{KD} sprouting defect. **a**, Intracellular aspartate levels upon acetate supplementation in control and CPT1A^{KD} ECs ($n = 3$). AU, arbitrary units. **b**, Intracellular levels of the indicated dNTPs upon acetate supplementation in control and CPT1A^{KD} ECs ($n = 3$). **c**, **d**, Percentage M+1 and M+2 labelling from [U-^{13}C]acetate in UMP (**c**) and UTP (**d**) in control and CPT1A^{KD} ECs ($n = 5$; P = NS). M denotes mass of the unlabelled metabolite. **e**, Total sprout length upon acetate supplementation in control and CPT1A^{KD} EC spheroids ($n = 3$). **f**, Total sprout length upon nucleoside mix supplementation in control and CPT1A^{KD} EC spheroids ($n = 4$). **g**, Total contribution of glutamine and glycolytic carbon to citrate in control and CPT1A^{KD} ECs ($n = 3$). **h**, Percentage M+2 labelled citrate from [U-^{13}C]glucose in control and CPT1A^{KD} ECs ($n = 3$). M denotes mass of the unlabelled metabolite. **i**, **j**, Percentage M+3 labelled malate (**i**) and aspartate (**j**) from [U-^{13}C]glucose in control and CPT1A^{KD} ECs ($n = 3$). M denotes mass of the unlabelled metabolite. Data are mean ± s.e.m. of n independent experiments. Statistical test: mixed models (**a**–**j**). NS, not significant. $***P < 0.001$, $****P < 0.0001$.

cycle were used for other purposes than deoxyribonucleotide synthesis. In fact, only fibroblasts incorporated similar amounts of fatty-acid-derived carbons into DNA to ECs (Fig. 6b).

Inhibition of CPT1A impairs angiogenesis

Pharmacological agents with FAO-blocking activity are being used clinically[11]. To explore whether pharmacological blockade of CPT1 reduced pathological angiogenesis, we used etomoxir, an irreversible inhibitor of mitochondrial long-chain FAO. Etomoxir reduced FAO and EC proliferation, but not migration *in vitro* (Extended Data Fig. 4a–c). *In vivo*, etomoxir induced similar retinal vascular defects as CPT1A deficiency in ECs (Extended Data Fig. 4d–g) and reduced pathological angiogenesis in a model of retinopathy of prematurity (Fig. 6c–e). The mechanistic model of our findings is represented in Fig. 6f.

Discussion

The novelty of this study relates to several observations. First, FAO is critical for vessel sprouting *in vivo*, and the extent of its function is greater than expected based on earlier reports[5]. Second, FAO affects proliferation, not migration, of ECs, which caused vessel sprouting defects. Unlike PFKFB3-driven glycolysis, which affects EC proliferation and migration[4], FAO selectively regulates EC proliferation, indicating that distinct metabolic pathways control distinct EC functions during vessel sprouting. Third, we discovered an unknown role of fatty-acid-derived carbons in *de novo* nucleotide synthesis for DNA replication in ECs. Our results indicate that FAO promotes *de novo* deoxyribonucleotide synthesis by providing carbons for the production of aspartate and glutamate. The fact that aspartate is a direct carbon donor for the synthesis of the pyrimidine nucleobase (Extended Data Fig. 3a) and that the pool of aspartate was reduced in CPT1A^{KD} ECs explains why pyrimidine dNTP levels were reduced. But aspartate and glutamate also serve as a nitrogen source for the production of the nucleobase of pyrimidines and purines. This explains why the synthesis of purine dNTPs was also impaired upon FAO blockade. Interestingly, of other primary and malignant cell types tested, only ECs and fibroblasts used substantial amounts of fatty acid carbons for DNA synthesis, even though certain cancer cell lines incorporated fatty-acid-derived carbons in the TCA cycle.

Fatty acids were irreplaceable for DNA synthesis, but not for protein and RNA synthesis. A possibility is that CPT1A^{KD} ECs maintained RNA levels by producing rNTPs via nucleotide salvage. Indeed, quiescent/hypo-proliferative cell types (CPT1A^{KD} ECs are hypo-proliferative) rely primarily on nucleotide salvage rather than on *de novo* nucleotide synthesis to synthesize RNA[12]. By contrast, proliferating cells switch on *de novo* synthesis of dNTPs to duplicate their DNA genome[13]. Support for the notion that ECs use nucleotide salvage primarily for RNA synthesis is provided by evidence that incorporation of carbons from hypoxanthine (a nucleotide salvage intermediate) was higher in RNA than DNA (Extended Data Fig. 3p). This may explain why FAO blockade affected dNTPs levels but not rNTP levels. A recent study on the role of pyruvate kinase isoenzymes documented that PKM1 expression reduced incorporation of ^{13}C label from [U-^{13}C]glucose and [U-^{13}C]glutamine in UMP and decreased dNTP levels and cell proliferation, yet did not alter rNTP levels[14], illustrating that the production of rNTPs and dNTPs is regulated differently.

Fourth, we identified that fatty-acid-derived carbons entering the TCA cycle are used for biomass production. This was unexpected since evidence in rapidly proliferating cancer cells indicated that glucose and glutamine are used as the primary carbon source for *de novo* deoxyribonucleotide synthesis[8,14]. The utilization of fatty acids for biomass production is, however, not unprofitable. Indeed, given that fatty acids are a much richer carbon source than glucose or glutamine, use of fatty acids is an efficient mechanism to generate nucleotides. Finally, pharmacological CPT1 blockade reduced pathological angiogenesis, which

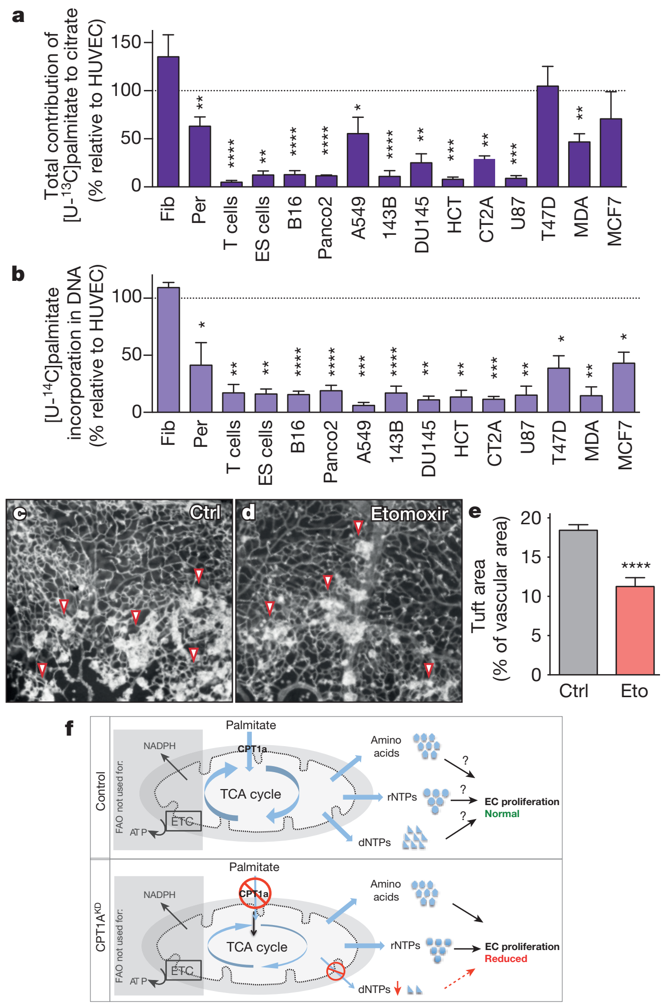


Figure 6 | Most other cell types do not use fatty acid carbons for dNTP synthesis, and inhibition of CPT1A impairs angiogenesis. **a**, Total contribution of [U-^{13}C]palmitate to citrate in various primary cells and cancer cell lines, expressed relative to the value in ECs ($n = 9$ for ECs, $n = 8$ for pericytes (Per), $n = 6$ for T cells, B16, A549 and 143B, $n = 5$ for Panco2, $n = 4$ for DU145, T47D and MDA, $n = 3$ for fibroblasts (Fib), ES cells, HCT, CT2A, U87 and MCF7). **b**, Contribution of [U-^{14}C]palmitate to DNA, expressed relative to the value in ECs ($n = 14$ for ECs, $n = 6$ for pericytes (Per), $n = 5$ for B16, Panco2 and A549, $n = 4$ for fibroblasts (Fib), T47D, 143B and T cells, $n = 3$ for HCT, MDA, DU145, U87, CT2A, MCF7 and ES cells). **c**, **d**, Representative images (original magnification, 10×) of retinal flat-mounts of retinopathy of prematurity (ROP) mice treated with vehicle (**c**) or etomoxir (Eto) (**d**). **e**, Vascular tuft area in control and etomoxir-treated pups ($n = 13$ for WT and 9 for etomoxir). **f**, Mechanistic model. Top, control ECs: uptake of palmitate and FAO in ECs are not essential for the production of ATP and NADPH (indicated in the shaded box), but fatty-acid-derived carbons are incorporated in amino acids, and in precursors of rNTPs and dNTPs; how critically these pathways regulate EC proliferation cannot be assessed in control cells (denoted by question mark). Bottom, CPT1A^{KD} ECs: silencing of CPT1A reveals, however, that decreasing FAO depletes dNTP pools, without affecting rNTP and protein synthesis, implying that fatty acids are irreplaceable for DNA synthesis; since *de novo* synthesis of dNTPs is critical for DNA replication, CPT1A silencing impairs EC proliferation. Data are mean ± s.e.m. of n independent experiments (**a**, **b**) or the total number of mice (of 4 litters) (**e**). Statistical test: two-sided t-test (**a**, **b**); mixed models (**e**). NS, not significant. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

may hint at an underappreciated therapeutic potential of lowering FAO for the inhibition of pathological angiogenesis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 June 2014; accepted 26 February 2015.

Published online 1 April 2015.

1. Potente, M., Gerhardt, H. & Carmeliet, P. Basic and therapeutic aspects of angiogenesis. *Cell* **146,** 873–887 (2011).
2. Ausprunk, D. H. & Folkman, J. Migration and proliferation of endothelial cells in preformed and newly formed blood vessels during tumor angiogenesis. *Microvasc. Res.* **14,** 53–65 (1977).
3. Welti, J., Loges, S., Dimmeler, S. & Carmeliet, P. Recent molecular discoveries in angiogenesis and antiangiogenic therapies in cancer. *J. Clin. Invest.* **123,** 3190–3200 (2013).
4. De Bock, K. *et al.* Role of PFKFB3-driven glycolysis in vessel sprouting. *Cell* **154,** 651–663 (2013).
5. De Bock, K., Georgiadou, M. & Carmeliet, P. Role of endothelial cell metabolism in vessel sprouting. *Cell Metab.* **18,** 634–647 (2013).
6. Colavitti, R. *et al.* Reactive oxygen species as downstream mediators of angiogenic signaling by vascular endothelial growth factor receptor-2/KDR. *J. Biol. Chem.* **277,** 3101–3108 (2002).
7. Fendt, S. M. *et al.* Reductive glutamine metabolism is a function of the α-ketoglutarate to citrate ratio in cells. *Nat. Commun.* **4,** 2236 (2013).
8. Vander Heiden, M. G. Exploiting tumor metabolism: challenges for clinical translation. *J. Clin. Invest.* **123,** 3648–3651 (2013).
9. Thompson, C. B. Wnt meets Warburg: another piece in the puzzle? *EMBO J.* **33,** 1420–1422 (2014).
10. Carracedo, A., Cantley, L. C. & Pandolfi, P. P. Cancer metabolism: fatty acid oxidation in the limelight. *Nature Rev. Cancer* **13,** 227–232 (2013).
11. Henry, T. D., Satran, D. & Jolicoeur, E. M. Treatment of refractory angina in patients not suitable for revascularization. *Nature Rev. Cardiol.* **11,** 78–95 (2014).
12. Morris, G. W., Iams, T. A., Slepchenko, K. G. & McKee, E. E. Origin of pyrimidine deoxyribonucleotide pools in perfused rat heart: implications for 3′-azido-3′-deoxythymidine-dependent cardiotoxicity. *Biochem. J.* **422,** 513–520 (2009).
13. Fairbanks, L. D., Bofill, M., Ruckemann, K. & Simmonds, H. A. Importance of ribonucleotide availability to proliferating T-lymphocytes from healthy humans. Disproportionate expansion of pyrimidine pools and contrasting effects of *de novo* synthesis inhibitors. *J. Biol. Chem.* **270,** 29682–29689 (1995).
14. Lunt, S. Y. *et al.* Pyruvate kinase isoform expression alters nucleotide synthesis to impact cell proliferation. *Mol. Cell* **57,** 95–107 (2015).

Acknowledgements We thank M. Vander Heiden and D. Tollervey for discussion, R. Adams for providing VE-cadherin(PAC)-Cre^{ERT2} mice, and S. Rodríguez-Arístegui for synthesis of etomoxir. S.S. is funded by the Institution of Research/Innovation (IWT); R.M., B.G., A.R.C. and J.G. by the Research Foundation Flanders (FWO); U.B. by a Marie Curie-IEF Fellowship; K.C.S.Q. by CAPES (Brazil) and G.B. by KU Leuven. The work of S.-M.F. is supported by Marie Curie CIG, FWO-OdysseusII, Concern Foundation, Bayer Healthcare Pharmaceuticals. The work of P.C. is supported by IUAP7/03, Methusalem funding (Flemish Government), FWO grants, Foundation Leducq Transatlantic Network (ARTEMIS), Foundation against Cancer, European Research Council (ERC) Advanced Research Grant (EU-ERC269073) and AXA Research grant. S.Y.L. was supported by the Department of Defense CDMRP Visionary Postdoctoral Award (W81XWH-12-1-0466). Views and opinions of, and endorsements by, the authors do not reflect those of the US Army or the Department of Defense. The authors thank the MSU LC-MS Core.

Author Contributions S.S., U.B., R.M., K.C.S.Q., G.B., I.E., A.Z., A.R.C., S.C., J.G., W.H., L.G., S. V., P.P.V.V., G.E., L.S., M.D., M.B., K.D.B., B.G., S.Y.L., S.-M.F. and P.C. contributed to the performance of the experiments and/or analysis of the data; H.G. provided advice; S.S., U.B., K.D.B., S.M.F. and P.C. designed the experiments; S.S., M.D., K.D.B., S.M.F. and P.C. wrote the paper; S.M.F. and P.C. conceptualized the metabolic analysis, and P.C. conceived and directed the study. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare competing financial interests: details are available in the online version of the paper. Correspondence and requests for materials should be addressed to P.C. (peter.carmeliet@vib-kuleuven.be) or S.-M.F. (sarah-maria.fendt@vib-kuleuven.be).

METHODS

Chemicals and reagents. Collagen type 1 (rat tail) was from Millipore or from PureCol (bovine) (Advanced Biomatrix, USA). The CPT1A inhibitor (+)-etomoxir sodium salt hydrate was purchased from CNIO Carlos III Therapies. Mitomycin C (MitoC), sodium palmitate, dimethyl sulfoxide (DMSO), *N*-acetyl-cysteine (NAC), sodium acetate, oligomycin, cycloheximide, cytidine, adenine, guanosine, methotrexate, carnitine, dNTP mix and tamoxifen were from Sigma-Aldrich (Bornem, Belgium). 5-fluorouracil (TEVA Pharma Belgium) was obtained from the pharmacy of the university hospital Leuven. Nucleoside mix was from Millipore (Belgium)[15]. Hoechst 33342 and L-homopropargylglycine (HPG) were from Molecular Probes and L-glutamine and penicillin/streptomycin were from Gibco (Invitrogen, Life Technologies, Ghent, Belgium). Uniformly labelled [U-^{13}C]potassium palmitate, [U-^{13}C]acetate, [U-^{13}C]glucose, [U-^{13}C]glutamine and [U-^{13}C]algal fatty acid mix were obtained from Cambridge isotope laboratories, Inc. [U-^{14}C]palmitate, [9,10-^{3}H]palmitate, [6-^{14}C]-D-glucose, [8-^{14}C]-hypoxanthine and [^{3}H]-thymidine were from Perkin Elmer.

Cell culture. Primary cells: human umbilical vein endothelial cells (HUVECs) were freshly isolated from different donors as described[16] (with approval of medical ethical committee KU Leuven/UZ Leuven and informed consent obtained from all subjects), regularly tested for mycoplasma and used between passage 1 and 5. HUVECs were cultured in M199 medium (1 mg ml^{-1} D-glucose) from Gibco (Invitrogen, Life Technologies, Ghent, Belgium) supplemented with 20% fetal bovine serum (FBS) Biochrom GmbH (Germany), 2 mM L-glutamine, 30 μg l^{-1} endothelial cell growth factor supplements (EGCS), 10 units ml^{-1} heparin (Sigma), 100 IU ml^{-1} penicillin and 100 μg ml^{-1} streptomycin or in endothelial basal medium (EBM-2) supplemented with endothelial growth medium SingleQuots (Promocell, Germany). In all experiments, HUVECs were always used as single-donor cultures. Human aortic ECs (HAECs) and adult human dermal blood microvascular ECs (HMVEC-dBlAd) were either from Clonetics, Lonza (Braine-l'Alleud, Belgium) or isolated from umbilical cords for arterial ECs or from placenta for microvascular ECs, and were cultured in the recommended medium supplemented with SingleQuots (EGM-2 or EGM-2 MV; Clonetics, Lonza, Braine-l'Alleud, Belgium). Human pericytes were isolated from human placenta using gentleMACS tissue dissociator (Miltenyi Biotec) and incubation with collagenase 0.2% (GIBCO) and DNase 80 U ml^{-1} (Life technologies) for 30 min at 37 °C followed by Percoll gradient 5 (GE Healthcare) centrifugation, or from human adipose tissue using collagenase 0.2% in 0.9% NaCl and 2 mM $CaCl_2$ for 1 h at 37 °C. Thereafter, the cell suspension was immunomagnetically sorted by using CD31 and CD146 antibodies conjugated with microbeads (Miltenyi Biotec). CD146^{+}CD31^{-} sorted cells were seeded in αMEM plus 10% FBS and 5 nM platelet derived growth factor-BB (PDGF-BB). Human T cells: peripheral blood mononuclear cells (PBMCs) were isolated from whole blood by Ficoll-Paque Plus (Amersham Biosciences) density gradient centrifugation. PBMCs were collected, washed with cold PBS and subsequently used for naive T-cell isolation. Naive T-cells were isolated with a pan T-cell isolation kit (MACS Myltenyi Biotec). After isolation, naive T cells were seeded in plates coated with anti-CD3 antibody (5 μg ml^{-1}), and anti-CD28 (0.5 μg ml^{-1}) and IL-2 (10 ng ml^{-1}) were added to the medium. All cell preparations were used in experiments as single donor cultures. Murine fibroblasts were freshly isolated from lungs of FVB mice and cultured and used between passage 1 and 5. Briefly, dissected lungs were chopped and digested in a solution of 2 mg ml^{-1} collagenase (Sigma blend L, Sigma-Aldrich) in RPMI media from Gibco (Invitrogen, Life Technologies, Ghent, Belgium) containing 2.5% fetal bovine serum (FBS), 10 mM HEPES, 2 mM glutamine, 100 IU ml^{-1} penicillin and 100 μg ml^{-1} streptomycin (5 ml per lung) for 1 h at 37 °C. Organoids were pelleted at 1,100 r.p.m., washed four times with digestion solution and then plated into standard tissue culture flasks in DMEM/F12 medium from Gibco (Invitrogen, Life Technologies, Ghent, Belgium) containing 5% fetal bovine serum (FBS), 2 mM glutamine, 100 IU ml^{-1} penicillin and 100 μg ml^{-1} streptomycin. The day after isolation, the medium was refreshed. Cells were then routinely maintained in 5% CO_2 and 95% air at 37 °C and the medium was replaced every second day. ES cells: G4 ES cells were routinely passaged every 2–3 days onto freshly prepared Mitomycin-C-arrested feeder layers, and were re-fed daily with fresh ES cell culture medium consisting of 500 ml Knockout DMEM medium (Life Technologies) supplemented with 2 mM L-glutamine (Life Technologies), 90 ml fetal bovine serum (Biochrom BmgH), MEM nonessential amino acids 100× (GIBCO/Invitrogen), 0.01 mM β-mercaptoethanol (Sigma-Aldrich), 1 mM sodium pyruvate (GIBCO/Invitrogen), penicillin/streptomycin 100× (Life Technologies), and 2,000 U ml^{-1} leukaemia inhibitory factor (Merck Millipore). Prior to testing whether ES cells use fatty acids for DNA synthesis they were replated at least 3 times on gelatinized dishes without any feeders. Doing so, all feeder cells were selectively removed from the cultures, resulting in 100% pure undifferentiated ES cell cultures. Tumour cell lines: murine melanoma B16 clone F10.9 (obtained from American Type Culture Collection, ATCC; LGC Standards, Molsheim Cedex, France) and Panco2 murine pancreatic carcinoma cells (provided by B. Wiedenmann, Charite¨, Berlin, Germany) were cultured in RPMI containing 10% FBS. MCF7 (human breast cancer), HCT116 (human colon carcinoma), MDA-MB-468 (human breast carcinoma cells), T47D (human ductal breast epithelial tumour cells), DU145 (human prostate cancer cells), 143B (human osteosarcoma cells) and A549 (human lung carcinoma cells) were cultured in DMEM containing 10% FBS, 2 mM glutamine, 100 IU ml^{-1} penicillin and 100 mg ml^{-1} streptomycin. CT2A murine glioblastoma cells and U87 human glioblastoma cells were cultured in DMEM containing 10% FBS and 2 mM glutamine (Invitrogen, Life Technologies, Ghent, Belgium) supplemented with conditioned medium obtained from CT2A or U87 cells.

Knock-down and overexpression strategies. RNA interference: CPT1A silencing was achieved using a validated pool of siRNA duplexes directed against human *CPT1A* (Trifekta Kit, IDT) and lipofectamine RNAi Max transfection reagent (Invitrogen, Belgium) according to the manufacturer's instructions. A scrambled siRNA was used as negative control. Lentiviral transductions: for overexpression of murine CPT1A, the cDNA was cloned in the pRRLsinPPT. CMV.MCS MM W prevector[17]. Of note, this increased FAO in some but not all EC isolations, suggesting that in such cases CPT1A abundance was not limiting the maximal capacity of FAO. However, whenever FAO was enhanced, sprout numbers and length of EC spheroids were increased. To generate shRNA vectors against *CPT1A*, *CPT1C*, or *ACADVL*, oligonucleotides were cloned into the pLKO-shRNA2 vector (No. PT4052-5; Clontech, Westburg BV, Leusden, The Netherlands) (for RNA sequences, see below). A nonsense scrambled shRNA sequence was used as a negative control. Production of lentiviruses by transfection into 293T cells was performed as described[18]. For transductions, a multiplicity of infection (MOI) of 20 was used in all shCPT1A experiments. For transductions with shACADVL and shCPT1C and for fluorescent ubiquitination-based cell cycle indicator (FUCCI; see below) and CPT1A overexpression, a MOI of 10 was used. Cells were transduced overnight and re-fed with fresh medium the next day. Transduced cells were used in functional assays at least 3 to 4 days after transduction. Of note, similar data were obtained using siRNA and two non-overlapping shRNAs directed against *CPT1A*.

***In vitro* assays.** Proliferation was quantified by incubating cells for 2–24 h with 1 μCi ml^{-1} [^{3}H]thymidine. Thereafter, cells were fixed with 100% ethanol for 15 min at 4 °C, precipitated with 10% trichloroacetic acid and lysed with 0.1 N NaOH. The amount of [^{3}H]thymidine incorporated into DNA was measured by scintillation counting. Scratch wound migration assay: a scratch wound was applied on confluent EC monolayers (pretreated with MitoC where indicated) using a 200 μl tip, 24 h after seeding (100,000 cells per well in 24-well plates). After scratch wounding and photography (T_0), the cultures were further incubated in fully supplemented EBM2 medium for 6–8 h (until near closure was reached in the control condition), and photographed again (T_x). Migration distance (gap area at T_0 minus gap area at T_x) was measured with NIH Image J software package and is expressed in arbitrary units. Modified Boyden chamber migration assays were performed using MitoC-treated (24 h treatment with 2 μg ml^{-1} MitoC) ECs. Therefore, 50,000 cells were seeded in fully supplemented EBM2 medium on 0.1% gelatine-coated transwells. Upon adherence, the transwells were washed and re-fed with medium containing only 0.1% FBS and transferred to bottom wells containing medium with 0.1% FBS (baseline) or 5% FBS (migration stimulus). Transwells were incubated for 16 h and processed and analysed for migrated cells as described[19]. Lamellipodia formation assay: the number and area of lamellipodia were measured on still photos of time-lapse recordings of sparsely seeded ECs, using the Axiovision morphometric analysis software (Carl Zeiss, Munich, Germany) with in-house developed macros. Lamellipodia area is expressed in per cent of total cell area. Directionality of migration was assessed by analysis of the total length of the migration track of a cell. The directionality value was then calculated by dividing the distance between the start and end location of the cell by the total track length; a higher directionality value denotes higher migration directionality. Experiments were performed with or without genetic shRNA *CPT1A* silencing. Spheroid capillary sprouting assay: ECs were incubated overnight in hanging drops in EGM-2 medium containing methylcellulose (20 volume% of a 1.2% solution of methylcellulose 4000 cP) (Sigma-Aldrich, Bornem, Belgium) to form spheroids. For mitotic inactivation, MitoC (2 μg ml^{-1}) was added to this medium. Spheroids were then embedded in collagen gel as described[20] and cultured for 20 h to induce sprouting. Compounds were added at the following concentrations during the collagen gel culture step, using corresponding vehicle concentrations as control: 20 mM sodium acetate, 5× nucleoside mix, 10 mM NAC, 1.2 μM oligomycin, 20 μg ml^{-1} cycloheximide, 50 μg ml^{-1} 5FU or 20 μM MTX. Cultures were fixed with 4% PFA at room temperature and imaged under bright field using a Motic AE 31 microscope (Motic Electric Group Co Ltd, Xiamen, China) or Leica DMI6000 microscope (Leica Microsystems, Mannheim, Germany). Analysis of the number of sprouts per spheroid and the total sprout length (cumulative length of primary sprouts and branches

per spheroid) was done on phase contrast images using the NIH Image J software package. To assess the number of viable versus dead cells, unfixed spheroids were stained with 500 μl of 1/500 dilution of TO-PRO3 (Invitrogen, Life Technologies, Ghent, Belgium) solution at 37 °C for 10 min, washed with PBS and imaged by confocal microscopy Zeiss LSM 510 Meta NLO or Zeiss LSM 780 confocal microscope (objectives: ×10 with NA 0.3, ×20 with NA 0.4) (Carl Zeiss, Munich, Germany)[21]. ATP measurement: analysis of total ATP levels was performed using a commercially available kit (ATPlite, PerkinElmer). Apoptosis: analysis of apoptosis was done by fluorescent staining for cleaved caspase 3 (Cell Signaling Technology, Bioke¨, Leiden, The Netherlands) in PFA-fixed monolayers.

RNA analysis. RNA expression analysis was performed by Taqman quantitative RT–PCR as described[22] using in-house-designed primers and probes or pre-made primer sets (Applied Biosystems, Carlsbad, CA and IDT, Belgium). Sequences or pre-made primer set ID numbers are available upon request. For comparison of gene expression between conditions, expression (normalized to HPRT endogenous control) is expressed relative to control condition. For comparison between different genes, absolute expression levels were determined based on respective cDNA standard curves, and levels are expressed as copies mRNA per 10^3 copies *HPRT* mRNA.

Immunoblot analysis. Protein extraction and immunoblot analysis were performed using a modified Laemmli sample buffer (125 mM Tris-HCl, pH 6.8 buffer containing 2% SDS and 10% glycerol)[22] or cell lysis buffer (Cell Signaling Technology) in the presence of protease and phosphatase inhibitors (Roche, Vilvoorde, Belgium). Lysates were separated by SDS–PAGE under reducing conditions, transferred to a nitrocellulose or PVDF membrane, and analysed by immunoblotting. Primary antibodies used were rabbit anti-CPT1A (No. 12252), rabbit anti-lamin A/C (No. 2032), rabbit anti-β-actin (No. 4970), anti-total AMPKα (No. 2532), rabbit anti-phospho-AMPKα Thr 172 (No. 2531), anti-p21 (No. 2947), anti-phospho-ATM (No. 5883), anti-total ATM (No. 2873) (Cell Signaling Technology, Bioke¨, Leiden, The Netherlands), and anti-total p53 (FL 393, Santa Cruz Biotechnology, The Netherlands). Autophagy was analysed via western blot determination of LC3 II/I ratio (No. 3868, Cell Signaling Technology) to reveal the on-state of autophagy (autophagic flux). Equal loading was verified by Ponceau Red staining and lamin A/C or β-actin as loading controls. Appropriate secondary antibodies were from Dako (Enschede, The Netherlands). Signal was detected using the ECL system (Amersham Biosciences, GE Healthcare, Diegem, Belgium) according to the manufacturer's instructions. Densitometric quantifications of bands were done with NIH Image J software.

Flow cytometry. Protein/RNA synthesis: protein synthesis was determined by incubation of cells with a methionine analogue (HPG, Invitrogen) for 24 h in full medium. RNA synthesis was determined by incubation of cells with a uridine analogue (5′-ethynyl uridine, EU, molecular probes). Cells were trypsinized, PFA fixed and incorporated HPG or 5′-ethynyl uridine was detected by a Click-IT reaction with Alexa fluor 647 according to the manufacturer's instructions. Alexa fluor 647 was excited with a 604-nm red laser and emission was recorded at 660 nm using a BD FACS Canto flow cytometer (Becton Dickinson Benelux NV, Erembodegem, Belgium). Intracellular ROS analysis: intracellular ROS levels were measured using the CM-DCF dye according to the manufacturer's instructions (Invitrogen, Belgium). For flow cytometry analysis the BD FACS Verse flow cytometer was used. For assessment of cell number, cells were seeded in a 24-well plate and trypsinized after 48 h growth. Cells were re-suspended in 150 μl of PBS and 100 μl of cell suspension was counted at a flow rate of 1 μl per second. Data were analysed using FlowJo analysis software.

Metabolism assays. Fatty acid oxidation: ECs were incubated in fully supplemented EBM2 medium with 100 μM unlabelled palmitate and 50 μM carnitine. Cells were incubated for 2 h in growth medium containing 2 $\mu Ci\ ml^{-1}$ [9,10-^{3}H]palmitate[23,24]. Thereafter, supernatant was transferred into glass vials sealed with rubber stoppers. ^3H_2O was captured in hanging wells containing a Whatman paper soaked with H_2O over a period of 48 h at 37 °C to reach saturation[25]. Radioactivity was determined by liquid scintillation counting. Glycolysis was measured analogously to fatty acid oxidation (cf supra) using 80 $mCi\ mmol^{-1}$ [5-^{3}H]-D-glucose (Perkin Elmer)[23]. ^{14}C-glucose oxidation: cells were incubated for 6 h in growth medium containing 100 $\mu Ci\ mmol^{-1}$ [6-^{14}C]-D-glucose. Thereafter, 250 μl of 2 M perchloric acid was added to each well to stop cellular metabolism and wells were immediately covered with a 1× hyamine hydroxide-saturated Whatman paper. Overnight absorption of $^{14}CO_2$ released during oxidation of glucose into the paper was performed at room temperature, and radioactivity in the paper was determined by liquid scintillation counting. ^{14}C-glutamine oxidation was performed similarly as glucose oxidation, except that we used 0.5 $\mu Ci\ ml^{-1}$ [U-^{14}C]glutamine as tracer. Palmitate or hypoxanthine mediated RNA and DNA synthesis was measured by the incorporation of ^{14}C into RNA or DNA using 100 $\mu Ci\ mmol^{-1}$ [U-^{14}C]palmitate or [8-^{14}C] hypoxanthine and was corrected for the total amount of RNA or DNA per sample. Total RNA and DNA were isolated using commercially available column-based DNA and RNA extraction kits (Qiagen) or using Trizol as an alternative extraction method for RNA or DNA. ATP coupled oxygen consumption: cells were seeded at 40,000 cells per well on Seahorse XF24 tissue culture plates (Seahorse Bioscience Europe, Copenhagen, Denmark). The measurement of oxygen consumption was performed at 10 min intervals (2 min mixing, 2 min recovery, 6 min measuring) for 3 h using the Seahorse XF24 analyser. For ATP-coupled oxygen consumption, measurements were performed before and after oligomycin (1.2 μM) treatment. Energy balance assessment: 1.5×10^6 cells were harvested in ice cold 0.4 M perchloric acid supplemented with 0.5 mM EDTA. pH was adjusted by adding 100 μl of 2 M K_2CO_3. 100 μl of the mixture was injected onto an Agilent 1260 HPLC equipped with a C18-Symmetry column (150 × 4.6 mm; 5 μm) (Waters), thermostated at 22.5 °C. Flow rate was kept constant at 1 $ml\ min^{-1}$. A linear gradient using solvent A (50 mM NaH_2PO_4, 4 mM tetrabutylammonium, adjusted to pH 5.0 using H_2SO_4) and solvent B (50 mM NaH_2PO_4, 4 mM tetrabutylammonium, 30% CH_3CN, adjusted to pH 5.0 using H_2SO_4) was accomplished as follows: 95% A for 2 min, from 2 to 25 min linear increase to 100% B, from 25 to 27 min isocratic at 100% B, from 27 to 29 min linear gradient to 95% A and finally from 29 to 35 min at 95% A. Detection of ATP, ADP and AMP occurred at 259 nm. GSSG/GSH ratio measurement: samples were collected in 300 μl 5% TCA (trichloro-acetic acid, Sigma). 50 μl was loaded onto an Ultimate 3000 UPLC (Thermo Scientific, Bremen, Germany) equipped with a Acquity UPLC HSS T3 column (catalogue no. 186003976; 2.1 × 5 mm; 1.8 μm particles; Waters) in line connected to a Q Exactive mass spectrometer (Thermo Fisher Scientific). A linear gradient was carried out using solvent A (0.05% formic acid) and solvent B (60% methanol, 0.05% formic acid). Practically, samples were loaded at 99% solvent A and from 10 to 12 min a ramp to 100% solvent B was carried out. From 15 to 16 min the column returned to 99% solvent A and the run was stopped at 21 min. Elution of GSH and GSSG occurred at 3 and 5.5 min, respectively (isocratic separation). Flow rate was constant at 250 $\mu l\ min^{-1}$ and the column temperature was kept constant at 37 °C. The mass spectrometer operated in targeted SIM mode following the ions m/z 311.11456 and 308.59499 (GSH and GSSG, respectively) using the ion 445.12003 as lock mass. The mass spectrometer ran in positive polarity, the source voltage was 3.0 kV, and the capillary temperature was set at 350 °C. Additional sheath gas flow was put at 35 and auxiliary gas flow rate at 10. Auxiliary gas heater temperature was put at 60 °C. AGC target was put at 100,000 ions with a maximum ion injection time of 200 ms, acquired at a resolution of 70,000. For the data analyses we manually integrated the peaks representing GSH and GSSG using the Thermo XCalibur Qual Browser software (Thermo Scientific) and data are represented as area of the respective GSH and GSSG peaks. Determination of ^{13}C-palmitate, glucose and glutamine incorporation in metabolites and total metabolite levels: for ^{13}C-carbon incorporation from palmitate in metabolites, cells were incubated for 48 h with labelled substrates (confirmation of steady state at that time was confirmed, see Extended Data Fig. 5). For ECs, [U-^{13}C]palmitate labelling was done in two ways: (1) '100% labelling', whereby all cold palmitate in M199 culture medium (120 μM) was replaced by 120 μM [U-^{13}C]palmitate using M199 medium, containing charcoal stripped serum (which does not contain any fatty acids); and (2) '50/50% labelling', whereby 100 μM [U-^{13}C] palmitate was added to the EGM2 culture medium containing 100 μM cold palmitate. Both types of labelling yielded similar data and were thus pooled. For comparison with cancer cells, only the 100% labelling strategy was used. Similar labelling methods were used for glucose (5.5 mM) and glutamine (2 mM). Labelling with the algal [U-^{13}C]fatty acid mix was performed by using 100% labelling; this fatty acid mix contained 50 μM palmitate. Metabolites for the subsequent mass spectrometry analysis were prepared by quenching the cells in liquid nitrogen followed by a cold two-phase methanol-water-chloroform extraction[7,26]. Phase separation was achieved by centrifugation at 4 °C and the methanol-water phase containing polar metabolites was separated and dried using a vacuum concentrator[24]. The dried metabolite samples were stored at −80 °C (refs 7, 26). Polar metabolites were derivatized for 90 min at 37 °C with 7.5 μl of 20 $mg\ ml^{-1}$ methoxyamine in pyridine and subsequently for 60 min at 60 °C with 15 μl of *N*-(tert-butyldimethylsilyl)-*N*-methyl-trifluoroacetamide, with 1% tert-butyldimethylchlorosilane[7,26]. Isotopomer distributions and metabolite levels were measured with a 7890A GC system (Agilent Technologies) combined with a 5975C Inert MS system (Agilent Technologies). One microlitre of sample was injected onto a DB35MS column in splitless mode using an inlet temperature of 270 °C (refs 7, 26). The carrier gas was helium with a flow rate of 1 $ml\ min^{-1}$. Upon injection, the GC oven was held at 100 °C for 3 min and then ramped to 300 °C with a gradient of 2.5 $°C\ min^{-1}$. The MS system was operated under electron impact ionization at 70 eV and a mass range of 100–650 amu (atomic mass units) was scanned. Isotopomer distributions were extracted from the raw ion chromatograms using a custom Matlab M-file, which applies consistent integration bounds and baseline correction to each ion[27]. In addition, we corrected for naturally occurring isotopes using the method of Fernandez *et al.*[28]. For relative metabolite levels, the total ion count was normalized to the internal standards norvaline and glutarate and to the protein

content[7,26]. To correct for enrichment dilution, we used previously reported methods[7,29]; that is, we divided the fractional contribution of a labelled metabolite of interest by the fractional contribution of its precursor (calculated by the formula below).

The total contribution of carbon was calculated using the following equation[7,29]:

$$\text{total contribution of carbon} = \sum_{i=0}^{n} i \times m_i \Big/ \left(n \sum_{i=0}^{n} m_i \right)$$

where n is the number of C atoms in the metabolite, i represents the different mass isotopomers and m refers to the abundance of a certain mass. Glycolytic carbon contribution was calculated based on [U-^{13}C]glucose labelling and label dilution in pyruvate[7]. For total metabolite levels, arbitrary units of the metabolite of interest were normalized to the protein content. A time-course experiment of the incorporation of [U-^{13}C]glucose, [U-^{13}C]glutamine and [U-^{13}C]palmitate in TCA intermediates demonstrated that the incorporation values reached a pseudo-isotopic steady state within experimental measurement precision (Extended Data Fig. 5). Determination of dNTP levels by RT–PCR: dNTP levels were determined by using a fluorescence-based PCR assay[30] using G1 sorted ECs, identified as CherryRed$^+$ Venus$^-$ cells upon transduction with a fluorescent, ubiquitination-based cell cycle indicator (FUCCI) construct[31]. Determination of ^{13}C-palmitate or ^{13}C-acetate incorporation in UMP and UTP: cells were labelled with [U-^{13}C]palmitate (100% labelling with 100 μM [U-^{13}C]palmitate; see above) or [U-^{13}C]acetate (20 mM supplementation with [U-^{13}C]acetate) for 48 h and were then collected in 500 μl ice-cold acetonitrile buffer (50% methanol, 30% acetonitrile and 20% water). Samples were spun for 5 min and supernatants were dried down and were then reconstituted in 50 μl of HPLC-grade water, vortexed, centrifuged, and transferred into HPLC vials. LC-MS/MS analysis was done on a Waters Xevo TQ-S mass spectrometry was coupled to an H-Class UPLC system. Metabolites were separated by polarity using a Supelco Ascentis Express C18 column (2.7 μm particle size, 5 cm × 2.1 mm). LC parameters are as follows: autosampler temperature, 5 °C; injection volume, 5 μl; column temperature, 50 °C; flow rate over 11 min: $t = 0$, 0.4 ml min^{-1}; $t = 2$, 0.3 ml min^{-1}; $t = 3$, 0.25 ml min^{-1}; $t = 5$, 0.15 ml min^{-1}; $t = 9$, 0.4 ml min^{-1}; $t = 11$, 0.4 ml min^{-1}. The LC solvents were solvent A: 10 mM tributylamine and 15 mM acetic acid in 97:3 water:methanol (pH 4.95); and solvent B: methanol. Elution from the column was performed over 11 min with the following gradient: $t = 0$, 0% B; $t = 1$, 0% B; $t = 2$, 20% B; $t = 3$, 20% B; $t = 5$, 55% B; $t = 8$, 95% B; $t = 8.5$, 95% B, $t = 9$, 0% B; $t = 11$, 0% B. Mass spectra were acquired using negative-mode electrospray ionization operating in multiple reaction monitoring (MRM) mode. The capillary voltage was 3,000 V, and cone voltage was 50 V. Nitrogen was used as cone gas and desolvation gas, with flow rates of 150 l h^{-1} and 600 l h^{-1}, respectively. The source temperature was 150 °C, and desolvation temperature was 500 °C. Argon was used as collision gas at a manifold pressure of 4.3×10^{-3} mbar. Collision energies and source cone potentials were optimized for each transition using Waters QuanOptimize software. Data were acquired and analysed using MassLynx 4.1 and QuanLynx software. Isotope labelling data was corrected for the natural abundance of different isotopes using IsoCor[32]. Determination of rNTP levels by LC-MS: rNTP was extracted with the same method as described for UTP and UMP. Additionally, ^{13}C-internal standards (generated based on fully labelled yeast extracts[33]) were spiked into the extraction solution. rNTP concentrations were determined with the same LC-MS method as described for UTP and UMP. All samples were normalized to the ^{13}C-internal standard and protein content.

Mouse models of ocular angiogenesis. Generation of EC-specific CPT1A conditional knockout mice: for generation of the *Cpt1a*$^{\text{lox/lox}}$ mice, *loxP* sites were introduced flanking a segment comprising exons 11 and 12 via homologous recombination in embryonic stem (ES) cells. The targeting vector contained from 5′ to 3′: a 3.5-kb 5′ homology flank, a 2-kb floxed mid-flank consisting of the *loxP* flanked exon 11–12 segment, a FRT flanked neomycine resistance (NEO) cassette, a 2.5-kb 3′ homology flank. Correctly targeted ES clones were identified by appropriate Southern blot and PCR analysis, and subjected to transient transfection with a Flp recombinase expression plasmid to remove the FRT-flanked NEO cassette. Correctly excised clones were used for morula aggregation, and resulting chimaeric animals were bred for germline transmission, yielding *Cpt1a*$^{\text{WT/lox}}$ and eventually *Cpt1a*$^{\text{lox/lox}}$ mice. For postnatal EC-specific loss of function studies, *VE-Cadherin(PAC)-Cre*^{ERT2} mice[34] (provided by R. Adams) were intercrossed with *Cpt1a*$^{\text{lox/lox}}$ mice and named CPT1A$^{\Delta\text{EC}}$. Correct Cre-mediated excision of the floxed *Cpt1a* segment in tamoxifen-treated CPT1A$^{\Delta\text{EC}}$ mice was confirmed via PCR analysis of genomic DNA using primers spanning the floxed region, by the appearance of a 300-bp band (Extended Data Fig. 1u). CPT1A$^{\Delta\text{EC}}$ mice were always compared with tamoxifen-treated Cre$^-$ *Cpt1a*$^{\text{lox/lox}}$ littermates (denoted as WT). Analysis of postnatal retinal angiogenesis: genetic endothelial-cell-specific inactivation of CPT1A in neonatal mice was achieved by intraperitoneal injection of 100 mg kg^{-1} tamoxifen solution (Sigma T5648; dissolved in 1:10 EtOH:corn oil) once daily from postnatal day (P) 1 to P4 in CPT1A$^{\Delta\text{EC}}$ mice. Pharmacological blockade of FAO in neonatal C57BL/6 WT mice (animal facility KU Leuven) was achieved by treatment with 30–35 mg kg^{-1} etomoxir from P2 to P4. For detection of cell proliferation, EdU (Invitrogen) was injected 2 h before dissection. At P5, pups were killed and eyes were enucleated, fixed with 4% PFA for 2 h at 4 °C and prepared for vascular analysis. As all animal treatments were done in baseline conditions, no randomization was required. Retinal whole mounts were prepared for vascular analysis as described[4]. Only litters for which the WT littermates (for CPT1A$^{\Delta\text{EC}}$ mice) or vehicle-treated littermates (for etomoxir treated mice) reached normal outgrowth and body weight at P5 were included for analysis. Radial vascular area, branching points, and number of filopodia were analysed on isolectin GS-IB4 stained retinas (see below) using the NIH Image J software package and Leica LASAF-MMAF morphometric analysis software (MetaMorph) (Leica Microsystems, Mannheim, Germany) with in-house developed macros. No statistical method was used to predetermine the sample size. Oxygen-induced retinopathy: oxygen-induced retinopathy (ROP) was induced by exposing C57BL/6 pups with their mother to 70% oxygen from P7 to P12 (ref. 35). Pups were then returned to normoxia and injected daily with 30 mg kg^{-1} etomoxir or vehicle. At P17, pups were killed and eyes were enucleated, fixed in 4% PFA and retinal flat-mounts were stained for isolectin B4, using previously published methods[4]. Mosaic images were captured using the inverted Leica DMI6000B epifluorescence microscope (Leica, Manheim, Germany) and analysis of the vascular tuft area was performed using NIH Image J software. For all mouse experiments, data analysis was done by experimenters blinded to the group allocation. All experimental animal procedures were approved by the Institutional Animal Care and Research Advisory Committee of the University of Leuven.

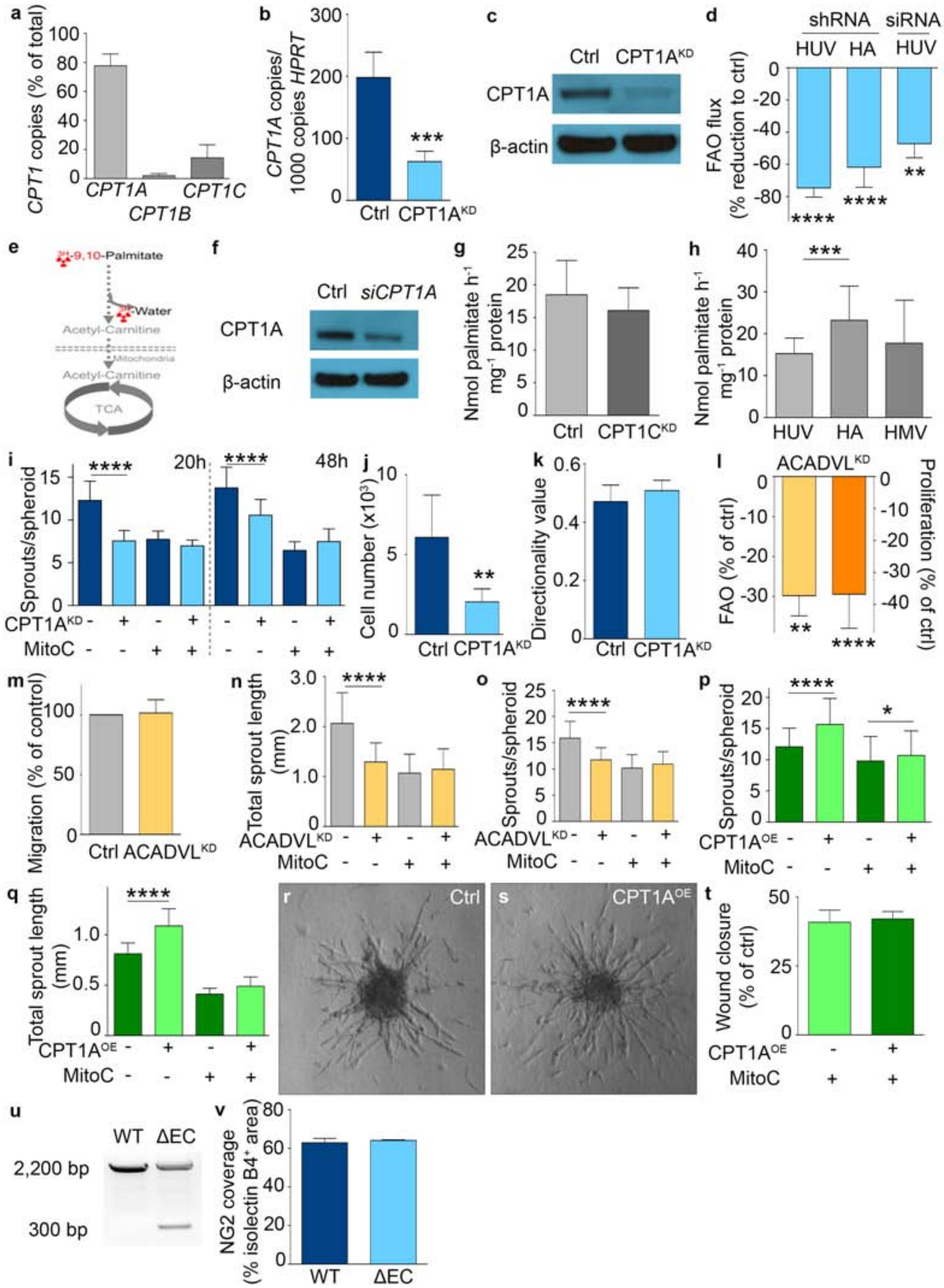
Immunohistochemistry. PFA-fixed cell cultures (4% PFA except for CPT1A staining where 2% PFA was used), spheroids (4% PFA) or whole-mount retinas (4% PFA) were subjected to immunofluorescence staining using the following isolectin conjugates or primary antibodies: isolectin GS-IB4-Alexa 488, isolectin GS-IB4-Alexa 568, isolectin GS-IB4-Alexa 647 (Molecular Probes), anti-CPT1A (Cell Signaling), anti-NG2 chondroitin sulfate proteoglycan (Chemicon), anti-Tomm20 (Abcam) and anti-collagen IV (Southern Biotech). Alexa-488, -568 or -633 conjugated secondary antibodies were used (Molecular Probes). EdU, EU and HPG staining was performed using a Click-IT assay with Alexa fluor 555 according to manufacturer's instructions.

Statistics. Data represent mean ± s.e.m. of pooled experiments unless otherwise stated. n values represent the number of independent experiments performed or the number of individual mice phenotyped. For each independent *in vitro* experiment, at least three technical replicates were used, except for the [U-^{13}C]palmitate incorporation analysis in cell lines (at least two technical replicates) (Fig. 6a). Statistical significance between groups was calculated using mixed model statistics (this test does not assume normality or equal variance) with experiment as random factor to correct for variation between umbilical donors (ECs were isolated from human umbilical cords) or mouse litters. SAS statistical software version 9.3 was used for statistical analysis. For comparison of ECs with other cell types, a one-sample t-test was used. Sample size for each experiment was not pre-determined. A P value <0.05 was considered statistically significant.

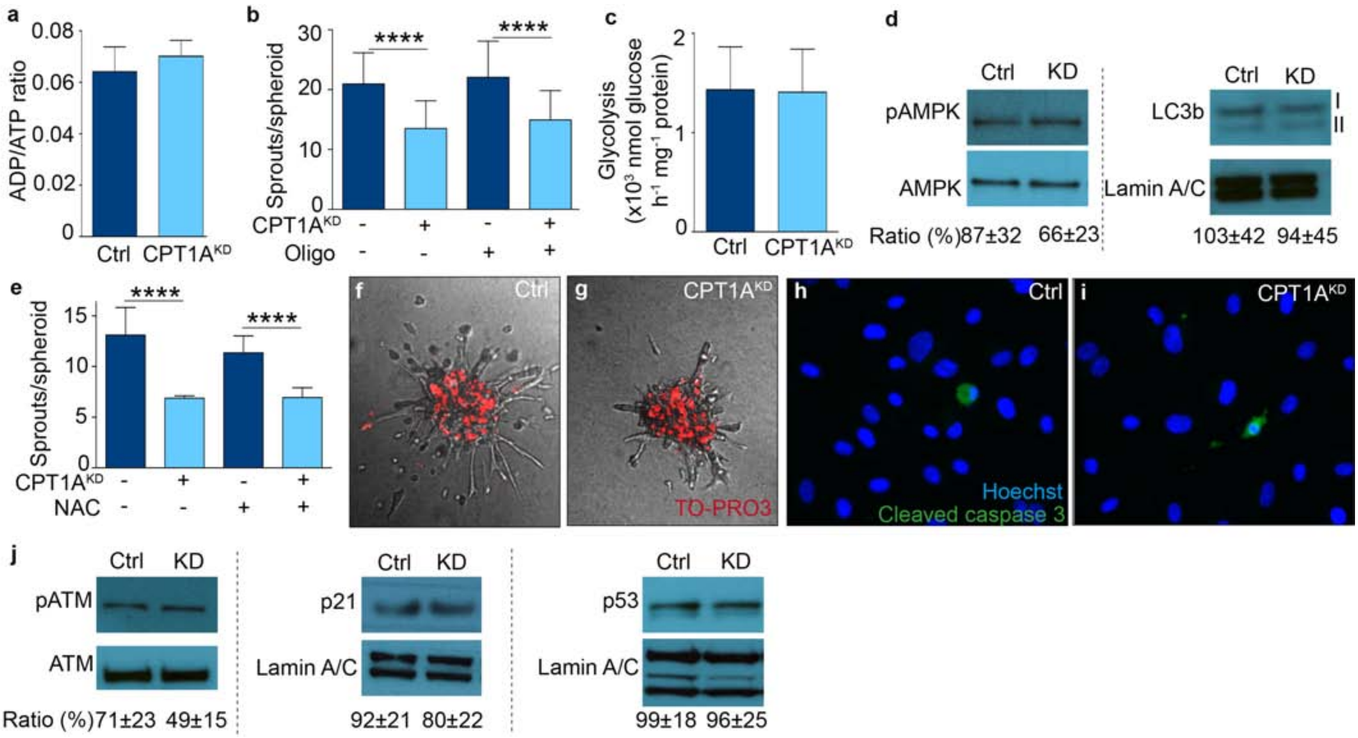
RNA sequences. shCPT1 RNA sequence 1, GCCATGAAGCTCTTAGACAAA; shCPT1 RNA sequence 2, CGATGTTACGACAGGTGGTTT; shACADVL RNA sequence, GCAGACATCTTCACGGTCTTT; siCPT1A RNA sequence (from IDT trifecta), AGCUCUUAGACAAAUC/AGAGAGAUAGAUUUGU.

15. Cheung, E. C. *et al.* TIGAR is required for efficient intestinal regeneration and tumorigenesis. *Dev. Cell* **25,** 463–477 (2013).
16. Jaffe, E. A., Nachman, R. L., Becker, C. G. & Minick, C. R. Culture of human endothelial cells derived from umbilical veins. Identification by morphologic and immunologic criteria. *J. Clin. Invest.* **52,** 2745–2756 (1973).
17. Michieli, P. *et al.* Targeting the tumor and its microenvironment by a dual-function decoy Met receptor. *Cancer Cell* **6,** 61–73 (2004).
18. Carlotti, F. *et al.* Lentiviral vectors efficiently transduce quiescent mature 3T3-L1 adipocytes. *Mol. Ther.* **9,** 209–217 (2004).
19. Geudens, I. *et al.* Role of delta-like-4/Notch in the formation and wiring of the lymphatic network in zebrafish. *Arterioscler. Thromb. Vasc. Biol.* **30,** 1695–1702 (2010).
20. Korff, T., Krauss, T. & Augustin, H. G. Three-dimensional spheroidal culture of cytotrophoblast cells mimics the phenotype and differentiation of cytotrophoblasts from normal and preeclamptic pregnancies. *Exp. Cell Res.* **297,** 415–423 (2004).
21. Schoors, S. *et al.* Partial and transient reduction of glycolysis by PFKFB3 blockade reduces pathological angiogenesis. *Cell Metab.* **19,** 37–48 (2014).
22. Carmeliet, P. *et al.* Synergism between vascular endothelial growth factor and placental growth factor contributes to angiogenesis and plasma extravasation in pathological conditions. *Nature Med.* **7,** 575–583 (2001).

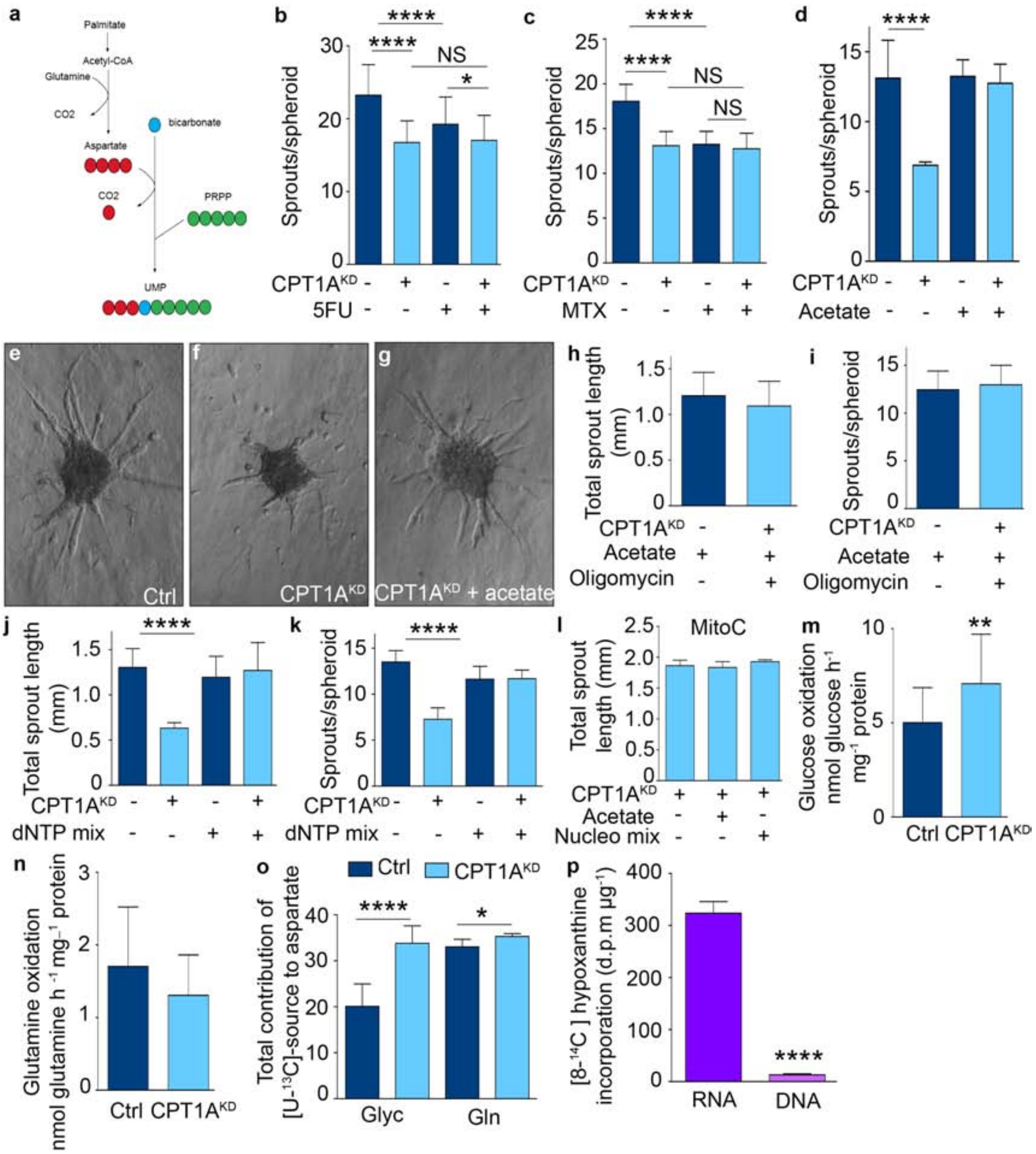
23. Dagher, Z., Ruderman, N., Tornheim, K. & Ido, Y. Acute regulation of fatty acid oxidation and amp-activated protein kinase in human umbilical vein endothelial cells. *Circ. Res.* **88,** 1276–1282 (2001).
24. Wang, R. *et al.* The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation. *Immunity* **35,** 871–882 (2011).
25. Aragone¨s, J. *et al.* Deficiency or inhibition of oxygen sensor Phd1 induces hypoxia tolerance by reprogramming basal metabolism. *Nature Genet.* **40,** 170–180 (2008).
26. Fendt, S. M. *et al.* Metformin decreases glucose oxidation and increases the dependency of prostate cancer cells on reductive glutamine metabolism. *Cancer Res.* **73,** 4429–4438 (2013).
27. Antoniewicz, M. R., Kelleher, J. K. & Stephanopoulos, G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.* **9,** 68–86 (2007).
28. Fernandez, C. A., Des Rosiers, C., Previs, S. F., David, F. & Brunengraber, H. Correction of 13^{C} mass isotopomer distributions for natural stable isotope abundance. *J. Mass Spectrom.* **31,** 255–262 (1996).
29. Nanchen, A., Fuhrer, T. & Sauer, U. Determination of metabolic flux ratios from ^{13}C-experiments and gas chromatography-mass spectrometry data: protocol and principles. *Methods Mol. Biol.* **358,** 177–197 (2007).
30. Wilson, P. M. *et al.* A novel fluorescence-based assay for the rapid detection and quantification of cellular deoxyribonucleoside triphosphates. *Nucleic Acids Res.* **39,** e112 (2011).
31. Sakaue-Sawano, A. *et al.* Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132,** 487–498 (2008).
32. Millard, P., Letisse, F., Sokol, S. & Portais, J. C. IsoCor: correcting MS data in isotope labeling experiments. *Bioinformatics* **28,** 1294–1296 (2012).
33. Büscher, J. M., Czernik, D., Ewald, J. C., Sauer, U. & Zamboni, N. Cross-platform comparison of methods for quantitative metabolomics of primary metabolism. *Anal. Chem.* **81,** 2135–2143 (2009).
34. Benedito, R. *et al.* The notch ligands Dll4 and Jagged1 have opposing effects on angiogenesis. *Cell* **137,** 1124–1135 (2009).
35. Scott, A. & Fruttiger, M. Oxygen-induced retinopathy: a model for vascular pathology in the retina. *Eye* **24,** 416–421 (2010).



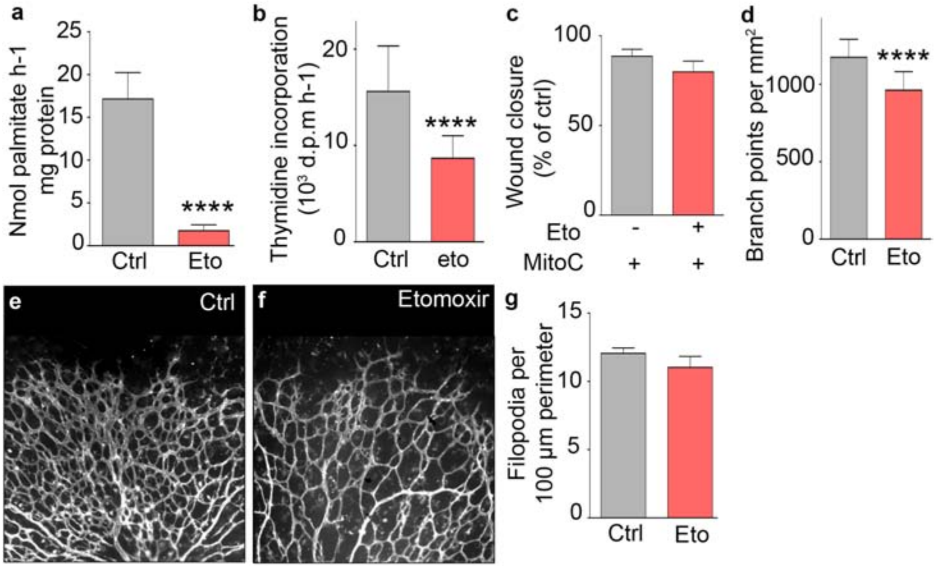
Extended Data Figure 1 | FAO regulates vessel sprouting. **a**, mRNA expression of CPT1 isoforms ($n = 3$). **b**, *CPT1A* mRNA levels upon *CPT1A* silencing (CPT1A^{KD}) ($n = 11$). **c**, Representative immunoblot of CPT1A for control and CPT1A^{KD} ECs. **d**, FAO flux upon CPT1A silencing in venous (HUV) and arterial (HA) ECs, or upon small interference RNA transfection in venous ECs (siRNA) ($n = 6$ for HUV shRNA, $n = 3$ for HA shRNA and HUV siRNA). **e**, Schematic representation of FAO measurement using [9,10-^{3}H] palmitate (reprinted from *Immunity* **35**, Wang, R. *et al.* The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation, 871–882 (2011), with permission from Elsevier; ref. 24). **f**, Representative immunoblot for CPT1A upon genetic silencing of *CPT1A* using siRNA (*siCPT1A*). **g**, FAO flux upon silencing of CPT1C (CPT1C^{KD}) ($n = 3$; P = NS). **h**, FAO flux levels in venous (HUV), arterial (HA) and microvascular (HMV) ECs ($n = 4$ for HUV versus HA and $n = 3$ for HUV versus HMV). **i**, Sprout number in control and CPT1A^{KD} EC spheroids with mitomycin C (MitoC) treatment as indicated ($n = 3$). **j**, Flow cytometry counting of viable control and CPT1A^{KD} ECs ($n = 3$). **k**, Analysis of random cell-motility tracks in control and CPT1A^{KD} ECs ($n = 4$; P = NS). **l**, FAO flux and proliferation upon silencing of ACADVL (ACADVL^{KD}) ($n = 3$ for each). **m**, Wound closure in control and ACADVL^{KD} ECs ($n = 3$; P = NS). **n**, **o**, Quantification of vessel sprouting in control and ACADVL^{KD} EC spheroids with MitoC treatment as indicated, total sprout length (**n**) and sprout numbers per spheroid (**o**) ($n = 5$). **p**, Sprout number in control and CPT1A^{OE} EC spheroids with MitoC treatment as indicated ($n = 3$). **q**, Total sprout length in control and CPT1A^{OE} EC spheroids treated with MitoC as indicated ($n = 5$). **r**, **s**, Representative phase (original magnification, 10× magnification) contrast images of control (**r**) and CPT1A^{OE} (**s**) EC spheroids. **t**, Scratch wound assay in control and CPT1A^{OE} ECs treated with MitoC as indicated ($n = 3$; P = NS). **u**, PCR analysis of genomic DNA from WT and CPT1A$^{\Delta EC}$ pups, confirming Cre-mediated recombination of the floxed *Cpt1a* allele as shown by the appearance of a 300-bp band. **v**, NG2^{+} area in neonatal vascular plexus of WT and CPT1A$^{\Delta EC}$ mice (3 litters, $n = 8$ pups for WT and 7 pups for CPT1A$^{\Delta EC}$; P = NS). Data are mean ± s.e.m. of n independent experiments (**a**, **b**, **d**, **g–q**, **t**) or the total number of mice (**v**). Statistical test: mixed models. NS, not significant. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.



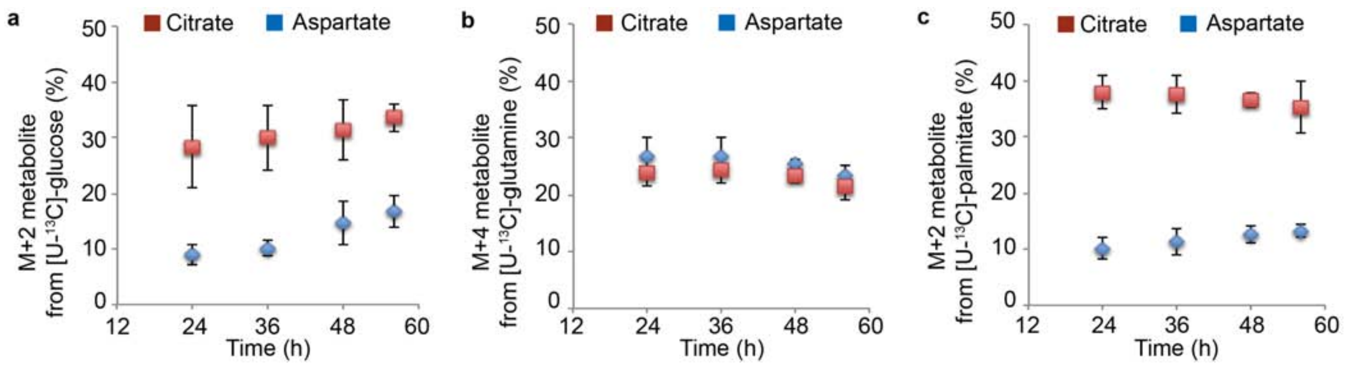
Extended Data Figure 2 | CPT1A silencing does not cause cellular distress. **a**, ADP/ATP ratio in control and CPT1A^{KD} ECs ($n = 3$; P = NS). **b**, Sprout number upon oligomycin treatment (oligo) in control and CPT1A^{KD} EC spheroids ($n = 3$). **c**, Glycolysis measurement in control and CPT1A^{KD} ECs ($n = 3$; P = NS). **d**, Representative immunoblot for phosphorylated AMPK (pAMPK) and total AMPK (AMPK) and for LC3b I and II in control and CPT1A^{KD} ECs. The ratio of the densitometrically quantified bands of pAMPK/AMPK and LC3b II/I is shown below the blots ($n = 3$; P = NS). **e**, Sprout number upon *N*-acetyl-cysteine (NAC) treatment in control and CPT1A^{KD} EC spheroids ($n = 3$). **f**, **g**, Representative images (original magnification, 10×) of EC spheroids upon staining for TO-PRO3 in control (**f**) and CPT1A^{KD} spheroids (**g**). **h**, **i**, Representative pictures (original magnification, 10×) of Hoechst/cleaved-caspase-3-stained control (**h**) and CPT1A^{KD} (**i**) ECs. **j**, Representative immunoblots showing the ratio of phosphorylated (pATM)/total-ATM (ATM), p21/lamin and p53/lamin in control and CPT1A^{KD} ECs. The ratios of the densitometrically quantified bands are shown below the blots ($n = 3$; P = NS). Data are mean ± s.e.m. of n independent experiments (**a**–**e**, **j**). Statistical test: mixed models. NS, not significant. ****$P < 0.0001$.



Extended Data Figure 3 | FAO is used for *de novo* nucleotide synthesis. **a**, Schematic representation of the different carbon sources used for *de novo* synthesis of UMP. Note that palmitate contributes three carbons to the nine carbon skeleton of UMP. PRPP, 5-phosphoribosyl-1-pyrophosphate. **b**, Sprout number upon 5-fluorouracil (5FU) treatment in control and CPT1A^{KD} EC spheroids ($n = 4$). **c**, Sprout number upon methotrexate (MTX) treatment in control and CPT1A^{KD} EC spheroids ($n = 4$). **d**, Sprout number upon acetate treatment in control and CPT1A^{KD} EC spheroids ($n = 3$). **e**–**g**, Representative images (original magnification, 10×) of EC spheroids upon acetate treatment. **h**, **i**, Rescue of the sprouting defect of CPT1A^{KD} spheroids by acetate was not affected by oligomycin treatment; panel **h**, total sprout length; panel **i**, sprout numbers/spheroid ($n = 3$; P = NS). **j**, **k**, Quantification of vessel sprouting using the EC spheroid model, showing that the reduction of total sprout length (**j**) and number of sprouts per spheroid (**k**) upon CPT1A silencing (CPT1A^{KD}) was rescued by supplementation with a dNTP mix ($n = 3$). **l**, Quantification of MitoC-treated EC spheroid sprouting upon acetate or nucleoside mix supplementation ($n = 3$; P = NS). **m**, Glucose oxidation in ECs, measured by $^{14}CO_2$ formation from [6-^{14}C]glucose in control and CPT1A^{KD} ECs ($n = 4$). **n**, Glutamine oxidation in ECs, measured by $^{14}CO_2$ formation from [U-^{14}C] glutamine in control and CPT1A^{KD} ECs ($n = 4$; P = NS). **o**, Total contribution of [U-^{13}C]glucose and [U-^{13}C]glutamine to aspartate in control and CPT1A^{KD} ECs ($n = 3$). **p**, [8-^{14}C]hypoxanthine incorporation in RNA and DNA in control ECs ($n = 3$). Data are mean ± s.e.m. of n independent experiments (**b**–**d**, **h**–**p**). Statistical test: mixed models. NS, not significant. $*P < 0.05$, $**P < 0.01$, $****P < 0.0001$.



Extended Data Figure 4 | Etomoxir reduces vessel sprouting. **a**, FAO flux upon etomoxir (Eto) treatment ($n = 6$). **b**, [^{3}H]thymidine incorporation upon etomoxir treatment ($n = 5$). **c**, Scratch wound assay using MitoC-treated ECs upon etomoxir (Eto) treatment ($n = 4$; $P =$ NS). **d**, Branch point quantification in the retinal vasculature of control (Ctrl) and etomoxir-treated (Eto) pups (8 litters, $n = 24$ pups for control and 16 for etomoxir treatment). **e**, **f**, Representative confocal images (original magnification, 10×) of retinal vessels stained for isolectin-B4 in control (**e**) and etomoxir (**f**) treated pups. **g**, Filopodia quantification in the retinal vascular front of control and etomoxir (Eto) treated pups (4 litters, $n = 11$ for WT and 9 for etomoxir; $P =$ NS). Data are mean ± s.e.m. of n independent experiments (**a**–**d**, **g**) or the total number of mice (**d**, **g**). Statistical test: mixed models. NS, not significant. ****$P < 0.0001$.



Extended Data Figure 5 | Analysis of steady state. Percentage M+2 or M+4 citrate and aspartate over different time points (24, 36, 48 and 52 h) after labelling with [U-^{13}C]glucose (**a**), [U-^{13}C]glutamine (**b**), or [U-^{13}C]palmitate (**c**). Data are mean ± s.d. of $n = 3$ independent experiments.

LETTER

doi:10.1038/nature14276

The comet-like composition of a protoplanetary disk as revealed by complex cyanides

Karin I. Öberg[1], Viviana V. Guzmán[1], Kenji Furuya[2], Chunhua Qi[1], Yuri Aikawa[3], Sean M. Andrews[1], Ryan Loomis[1] & David J. Wilner[1]

Observations of comets and asteroids show that the solar nebula that spawned our planetary system was rich in water and organic molecules. Bombardment brought these organics to the young Earth's surface[1]. Unlike asteroids, comets preserve a nearly pristine record of the solar nebula composition. The presence of cyanides in comets, including 0.01 per cent of methyl cyanide (CH_3CN) with respect to water, is of special interest because of the importance of C–N bonds for abiotic amino acid synthesis[2]. Comet-like compositions of simple and complex volatiles are found in protostars, and can readily be explained by a combination of gas-phase chemistry (to form, for example, HCN) and an active ice-phase chemistry on grain surfaces that advances complexity[3]. Simple volatiles, including water and HCN, have been detected previously in solar nebula analogues, indicating that they survive disk formation or are re-formed *in situ*[4–7]. It has hitherto been unclear whether the same holds for more complex organic molecules outside the solar nebula, given that recent observations show a marked change in the chemistry at the boundary between nascent envelopes and young disks due to accretion shocks[8]. Here we report the detection of the complex cyanides CH_3CN and HC_3N (and HCN) in the protoplanetary disk around the young star MWC 480. We find that the abundance ratios of these nitrogen-bearing organics in the gas phase are similar to those in comets, which suggests an even higher relative abundance of complex cyanides in the disk ice. This implies that complex organics accompany simpler volatiles in protoplanetary disks, and that the rich organic chemistry of our solar nebula was not unique.

MWC 480 is a Herbig Ae star with an estimated stellar mass of 1.8 solar masses ($M_\odot$)[9] in the Taurus star-forming region at a distance of 140 pc. The star is surrounded by a $0.18 \pm 0.1\ M_\odot$ protoplanetary disk; this is an order of magnitude more massive than the $0.01\ M_\odot$ minimum-mass solar nebula, the lowest possible mass of the solar nebula that could have produced the Solar System[10,11]. Compared to disks around solar-type stars, the MWC 480 disk is 2–3 times warmer at a given radius[12,13] and is exposed to levels of ultraviolet radiation orders of magnitude higher. Despite these environmental differences, the composition and abundance of volatiles in the MWC 480 disk appear largely similar to those in disks around solar-type stars, except for a lower abundance of cold (temperature $T < 20$ K) chemistry tracers in the outer disk[14,15]. The inner disk chemistry of MWC 480 has not been studied, but Herbig Ae and T Tauri disks are observed to have different volatile compositions close to their stars[16].

Using the Atacama Large Millimeter/submillimeter Array (ALMA), we detected two rotational emission lines of CH_3CN from the MWC 480 protoplanetary disk (the 14_0–13_0 line at 5σ, and the 14_1–13_1 line at 3.5σ, with each energy level characterized by the quantum numbers J and K, the total angular momentum and projection of the angular momentum along the molecular symmetry axis, respectively). We also detected emission lines from the N-bearing carbon chain HC_3N, and from the ^{13}C isotopologue of HCN. We targeted the ^{13}C isotopologue rather than the more abundant $H^{12}CN$ because the latter is optically thick and therefore a poor tracer of the HCN abundance. Figure 1 shows the spatially resolved line detections together with a dust continuum emission map,

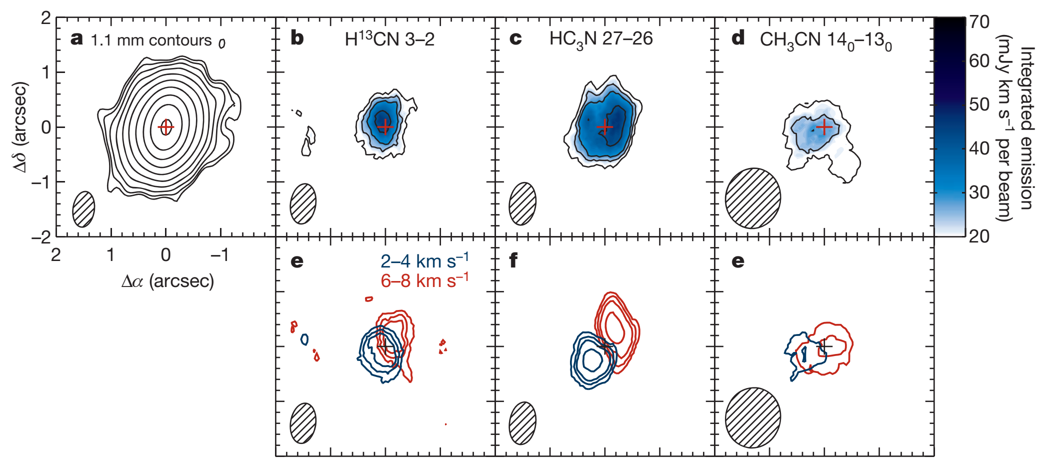


Figure 1 | ALMA detections of simple and complex cyanides in the MWC 480 protoplanetary disk. **a**, 1.1 mm emission (black contours are $3\sigma + \sigma \times 2^{[1,2,\ldots]}$). **b–d**, Integrated emission of $H^{13}CN$ (**b**), HC_3N (**c**) and CH_3CN (**d**) lines (colour: see colour scale on the right). Black contours are $[3,4,5,7,10]\sigma$. **e–g**, As **b–d**, but for 2 km s^{-1} velocity bins around the source mean velocity, displaying the disk rotation. Positions are relative to the continuum phase centre (marked with a cross) at right ascension (α) 04 h 58 min 45.94 s and declination (δ) +29° 50′ 38.4″. The synthesized beam is shown in the bottom left corner of each panel.

[1]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [2]Leiden Observatory, Leiden University, PO Box 9513, 2300 CA Leiden, The Netherlands. [3]Kobe University, 1-1 Rokkodaicho, Nada Ward, Kobe, Hyogo Prefecture 657-0013, Japan.

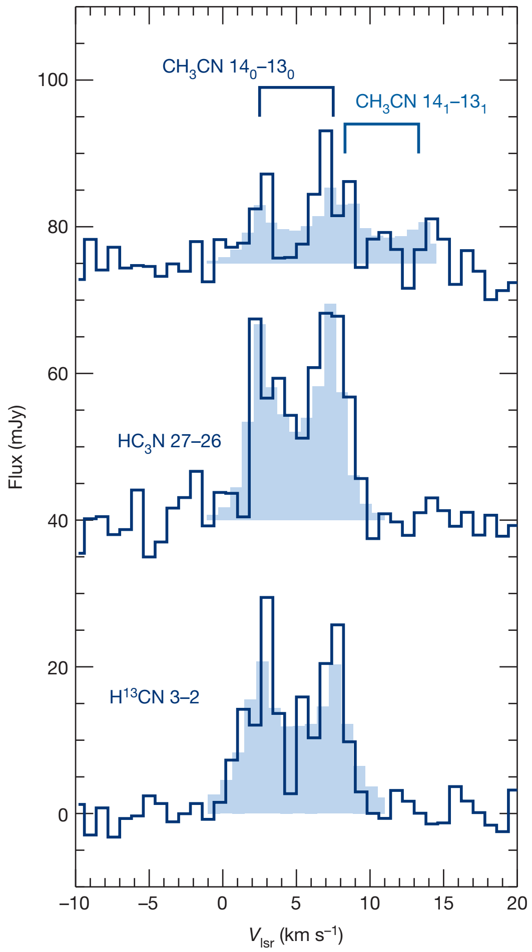


Figure 2 | Spectra of detected cyanides in the MWC 480 protoplanetary disk. The observed spectra (contours) of $H^{13}CN$, HC_3N and CH_3CN are extracted from ALMA spectral-image data cubes and are shown as functions of the local standard of rest velocity, V_{lsr}. The synthetic spectra (light blue shaded histograms) are based on the best-fit disk abundance models in Fig. 3. The CH_3CN spectrum contains two partially overlapping lines identified with the 14_0–13_0 and 14_1–13_1 transitions. The spectra were extracted from the spectral image cubes using a Keplerian mask to maximize the signal-to-noise ratio.

demonstrating the spatial coincidence between CH_3CN, HC_3N, HCN and dust emission from the disk. The angular resolution is 0.4″–0.6″, corresponding to 50–70 astronomical units (1 AU is the distance from the Earth to the Sun). The emission is also spectrally resolved, which can be used to probe smaller spatial scales. Figure 1 shows the velocity gradient across the disk that arises from Keplerian rotation in all three lines. Figure 2 shows the spectra of the three lines; each displays the double-peaked structure typical of a rotating disk. Table 1 lists the integrated line fluxes.

We use the spectrally and spatially resolved line emission to constrain the radial profiles of molecular column density based on parametric abundance models defined with respect to the adopted density and temperature structure of the MWC 480 disk (Methods). Figures 2 and 3a–c show the synthetic line spectra and maps from the best-fit models, demonstrating the good match between models and data. Figure 3d–f shows the best-fit radial column density profile, together with all profiles consistent with the data within 3σ, and abundances at 30 and 100 AU from the central star are reported in Table 1. The best-fit profiles have different slopes for the different molecules. The $H^{13}CN$ column density decreases with radius, which is consistent with predictions from disk chemistry models[17]. The increasing column density with radius out to 100 AU of HC_3N, effectively a ring, is not predicted by models[18], indicating that disk chemistry models are incomplete for HC_3N. The CH_3CN emission is best reproduced with a flat profile, but other profiles cannot be excluded.

The absolute abundances depend on the assumed disk density structure, and we therefore compare the complex cyanides in the MWC 480 disk to cometary and to protostellar compositions using HCN as a reference species, similar to the practice in cometary studies[19]. We calculate abundances of CH_3CN and HC_3N with respect to HCN at 30 AU, the smallest disk radius accessible by the ALMA observations, and at 100 AU, the outer boundary of the cyanide emission maps. Accounting for the higher luminosity of MWC 480 compared to the young Sun, this radial range in the MWC 480 disk corresponds to the comet-forming zone of 10–30 AU in the solar nebula[19]. Assuming a standard HCN/$H^{13}CN$ ratio of 70, the best-fit HC_3N and CH_3CN abundances with respect to HCN are 0.4 and 0.05 at 30 AU, and 5 and 0.2 at 100 AU, respectively. The CH_3CN/HCN abundance ratios are robust to model assumptions to within factors of a few, while the HC_3N/HCN abundance ratio may be overestimated by an order of magnitude when using our simple abundance model (Methods). Conservatively, both the CH_3CN and HC_3N abundances with respect to HCN are thus ~5% at 30 AU and ~20% at 100 AU. A typical comet contains 10% of CH_3CN and HC_3N with respect to HCN (ref. 19). The MWC 480 gas-phase cyanide composition at both 30 and 100 AU is thus cometary within the observational and model uncertainties.

The relationship between gas-phase abundance ratios and the abundance ratios in ices, the main reservoirs of volatiles in disks[17,18,20], depends on both desorption characteristics and chemistry (Methods). HCN, HC_3N and CH_3CN are characterized by similar freeze-out and desorption kinetics, but different chemical pathways. In particular, the existence of efficient grain surface formation pathways to CH_3CN enhances CH_3CN with respect to the other cyanides in the ice mantles. The scale of this enhancement factor varies among models, but it is at least one order of magnitude (Methods). This results in an expected minimum CH_3CN/HCN ice ratio of 0.5 at 30 AU in the MWC 480 disk, considerably higher than that found in the typical Solar System comet.

The CH_3CN/HCN and HC_3N/HCN ratios in the MWC 480 disk are also high when compared to protostars. The HCN/HC_3N/CH_3CN ratio is 1/0.01/0.08 towards the solar-type protostellar binary IRAS 16298-2422[21], and similar abundance ratios are found towards more massive systems. The MWC 480 cyanide composition is thus difficult to explain by inheritance alone, as has been suggested for H_2O, for example (ref. 7). Rather, the observed high disk abundances probably reflect an efficient disk chemistry that readily converts a large portion of the carbon originally in CO and other small molecules into more complex organics[22,23] during the first million years of the disk life time.

Table 1 | Molecular data

Molecule	QN	Rest frequency (GHz)	E_u (K)	Integrated flux (mJy km s^{-1} per beam)	x^*_{30AU} ($10^{-13} n_H^{-1}$)	x^*_{100AU} ($10^{-13} n_H^{-1}$)
$H^{13}CN$	$J = 3$–2	259.0118	24.9	47 ± 8	1.6 [1.3–2.0]	1.6 [1.3–2.0]
HC_3N	$J = 27$–26	245.6063	165	66 ± 8	43 [36–54]	480 [400–600]
CH_3CN	14_0–13_0	257.5274	92.7	30 ± 6	4.8 [2.3–7.5]	16 [6–55]
CH_3CN	14_1–13_1	257.5224	100	23 ± 6		

QN, quantum numbers of the transition; E_u, energy of the upper level.
*Best-fit abundances at 30 and 100 AU assuming a vertically constant abundance. The 3σ abundance range is in square brackets.

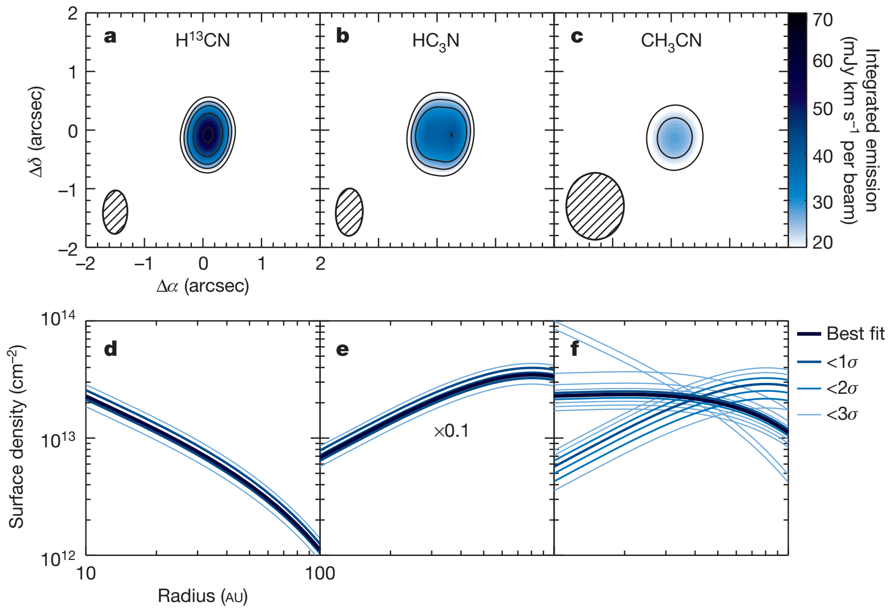


Figure 3 | Models of cyanide emission and radial distributions in the MWC 480 disk. **a**–**c**, Synthetic ALMA observations of the integrated line emission flux without added noise based on the best-fit radial column density profiles of $H^{13}CN$, HC_3N, and CH_3CN (colour: see colour scale on the right). Black contours are as in Fig. 1. **d**–**f**, Best-fit radial column density profiles (thick dark blue line), shown together with all column density profiles that are consistent with data within 1, 2 and 3σ confidence intervals. Increasing deviations from the best fit are shown with successively lighter shades and thinner lines.

This early, efficient complex chemistry in protoplanetary disks affects the surface conditions of rocky (exo-)planets. In the 'grand tack' model of early Solar System dynamics[24], the coupled migration of Jupiter and Saturn in the gas-rich disk from which they formed caused scattering of volatile-rich (and thus complex-organic-rich) planetesimals inwards, mixing with the asteroid belt, and outwards, producing the present-day Kuiper belt and comets[25]. While comets are expected (and observed) to conserve most of their original compositions due to low temperatures and thus very long desorption and chemical timescales[19], asteroid-belt bodies are more processed. A combination of evaporation and chemistry probably resulted in both a net loss of volatiles and in a relative enhancement of complex organics with respect to H_2O over the subsequent tens of Myr (ref. 26). A second instability 100 Myr later resulted in a heavy bombardment of the Earth by mainly asteroid-belt bodies, including the icy bodies originating in the outer Solar System[27]. The frequency of these instabilities in exoplanetary systems is unknown, but the large numbers of 'hot Jupiters' and 'super-Earths' too close to their host stars (within 0.1 AU) to be explained by *in situ* formation reveal that major planetary migrations are common. When these migrations cause bombardment of icy bodies, the icy bodies in question are probably organic-rich.

The high ratio of complex to simple cyanides in the MWC 480 disk implies that the rich organic composition of comets is not unique to our Solar System, and could be commonplace. Laboratory experiments have shown that the same ice chemistry that produces CH_3CN, that is, photo-processing of interstellar ice analogues, also produces simple sugars and amino acids[28,29]. This suggests that the early surface conditions of Earth, set by comet and asteroid bombardment, may be common for young rocky planets, and that conditions favourable to an even richer chemistry may be ubiquitous.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 November 2014; accepted 30 January 2015.

1. Hartogh, P. *et al.* Ocean-like water in the Jupiter-family comet 103P/Hartley 2. *Nature* **478,** 218–220 (2011).
2. Goldman, N., Reed, E. J., Fried, L. E., William Kuo, I.-F. & Maiti, A. Synthesis of glycine-containing complexes in impacts of comets on early Earth. *Nature Chem.* **2,** 949–954 (2010).
3. Herbst, E. & van Dishoeck, E. F. Complex organic interstellar molecules. *Annu. Rev. Astron. Astrophys.* **47,** 427–480 (2009).
4. Dutrey, A., Guilloteau, S. & Guelin, M. Chemistry of protosolar-like nebulae: the molecular content of the DM Tau and GG Tau disks. *Astron. Astrophys.* **317,** L55–L58 (1997).
5. Carr, J. S. & Najita, J. R. Organic molecules and water in the planet formation region of young circumstellar disks. *Science* **319,** 1504–1506 (2008).
6. Hogerheijde, M. R. *et al.* Detection of the water reservoir in a forming planetary system. *Science* **334,** 338–340 (2011).
7. Cleeves, L. I. *et al.* The ancient heritage of water ice in the solar system. *Science* **345,** 1590–1593 (2014).
8. Sakai, N. *et al.* Change in the chemical composition of infalling gas forming a disk around a protostar. *Nature* **507,** 78–80 (2014).
9. Simon, M., Dutrey, A. & Guilloteau, S. Dynamical masses of T Tauri stars and calibration of pre-main-sequence evolution. *Astrophys. J.* **545,** 1034–1043 (2000).
10. Guilloteau, S., Dutrey, A., Piétu, V. & Boehler, Y. A dual-frequency sub-arcsecond study of proto-planetary disks at mm wavelengths: first evidence for radial variations of the dust properties. *Astron. Astrophys.* **529,** A105 (2011).
11. Takami, M. *et al.* Surface geometry of protoplanetary disks inferred from near-infrared imaging polarimetry. *Astrophys. J.* **795,** 71 (2014).
12. Chiang, E. I. & Goldreich, P. Spectral energy distributions of T Tauri stars with passive circumstellar disks. *Astrophys. J.* **490,** 368–376 (1997).
13. D'Alessio, P., Calvet, N., Hartmann, L., Muzerolle, J. & Sitko, M. in *Star Formation at High Angular Resolution* (eds Burton, M. G., Jayawardhana, R. & Bourke, T. L.) 403–410 (IAU Symp. Vol. 221, 2004).
14. Dutrey, A. *et al.* Chemistry in disks. I. Deep search for N_2H^+ in the protoplanetary disks around LkCa 15, MWC 480, and DM Tauri. *Astron. Astrophys.* **464,** 615–623 (2007).
15. Öberg, K. I. *et al.* The disk imaging survey of chemistry with SMA. I. Taurus protoplanetary disk data. *Astrophys. J.* **720,** 480–493 (2010).
16. Pontoppidan, K. M. *et al.* A Spitzer survey of mid-infrared molecular emission from protoplanetary disks. I. Detection rates. *Astrophys. J.* **720,** 887–903 (2010).
17. Walsh, C., Millar, T. J. & Nomura, H. Chemical processes in protoplanetary disks. *Astrophys. J.* **722,** 1607–1623 (2010).
18. Walsh, C. *et al.* Complex organic molecules in protoplanetary disks. *Astron. Astrophys.* **563,** A33 (2014).
19. Mumma, M. J. & Charnley, S. B. The chemical composition of comets: emerging taxonomies and natal heritage. *Annu. Rev. Astron. Astrophys.* **49,** 471–524 (2011).
20. Semenov, D. & Wiebe, D. Chemical evolution of turbulent protoplanetary disks and the solar nebula. *Astrophys. J. Suppl. Ser.* **196,** 25 (2011).
21. van Dishoeck, E. F., Blake, G. A., Jansen, D. J. & Groesbeck, T. D. Molecular abundances and low-mass star formation. II. Organic and deuterated species toward IRAS 16293–2422. *Astrophys. J.* **447,** 760–782 (1995).
22. Ciesla, F. J. & Sandford, S. A. Organic synthesis via irradiation and warming of ice grains in the solar nebula. *Science* **336,** 452–454 (2012).

23. Favre, C., Cleeves, L. I., Bergin, E. A., Qi, C. & Blake, G. A. A significantly low CO abundance toward the TW Hya protoplanetary disk: a path to active carbon chemistry? *Astrophys. J.* **776,** L38 (2013).
24. Tsiganis, K., Gomes, R., Morbidelli, A. & Levison, H. F. Origin of the orbital architecture of the giant planets of the Solar System. *Nature* **435,** 459–461 (2005).
25. Walsh, K. J., Morbidelli, A., Raymond, S. N., O'Brien, D. P. & Mandell, A. M. A low mass for Mars from Jupiter's early gas-driven migration. *Nature* **475,** 206–209 (2011).
26. Pizzarello, S. Catalytic syntheses of amino acids and their significance for nebular and planetary chemistry. *Meteorit. Planet. Sci.* **47,** 1291–1296 (2012).
27. O'Brien, D. P., Walsh, K. J., Morbidelli, A., Raymond, S. N. & Mandell, A. M. Water delivery and giant impacts in the Grand Tack scenario. *Icarus* **239,** 74–84 (2014).
28. Öberg, K. I., Garrod, R. T., van Dishoeck, E. F. & Linnartz, H. Formation rates of complex organics in UV irradiated CH_3OH-rich ices. I. Experiments. *Astron. Astrophys.* **504,** 891–913 (2009).
29. Muñoz Caro, G. M. *et al.* Amino acids from ultraviolet irradiation of interstellar ice analogues. *Nature* **416,** 403–406 (2002).

Acknowledgements We acknowledge comments from E. van Dishoeck. This Letter makes use of ALMA data. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada) and NSC and ASIAA (Taiwan), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. The National Radio Astronomy Observatory is a facility of the NSF operated under cooperative agreement by Associated Universities, Inc. K.I.Ö. acknowledges A. Leroy and the NAASC for assistance with calibration and imaging, and also acknowledges funding from the Simons Collaboration on the Origins of Life (SCOL), the Alfred P. Sloan Foundation, and the David and Lucile Packard Foundation. D.J.W. acknowledges funding from NASA Origins of Solar Systems (grant no. NNX11AK63).

Author Contributions K.I.Ö. led the overall project, reduced the data, assisted by V.V.G. and R.L., and wrote the manuscript with revisions from S.M.A. and D.J.W. V.V.G., assisted by C.Q., performed the parametric modelling and abundance extraction. K.F. performed the astrochemical modelling, and interpreted the results with Y.A. All authors contributed to discussions of the results and commented on the manuscript.

Author Information The ALMA program number for the presented data is 2013.1.00226. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.I.Ö. (koberg@cfa.harvard.edu).

METHODS

ALMA observations and data reduction. MWC 480 was observed with the Atacama Large Millimeter/submillimeter Array (ALMA) as a part of the Cycle 2 proposal 2013.1.00226. Observations were carried out on 2014 June 15, with 33 antennas and baselines of 18–650 m (15–555 kλ). The total on-source integration time was 22 min. The nearby quasar J0510+1800 was used for phase and gain, bandpass and absolute flux calibration (J0510+1800 was itself calibrated to an absolute flux of 1.50 ± 0.08 Jy on 2014 June 29). The correlator was configured to observe 14 spectral windows (SPWs) with resolution $\delta\nu = 61.04$ kHz. The $H^{13}CN$ 3–2 line was placed in SPW 1, HC_3N 27–26 in SPW 9, and CH_3CN 14_0–13_0 and 14_1–13_1 in SPW 5, each 60–120 MHz wide and centred on 258.983, 245.622 and 257.498 GHz, respectively.

The visibility data were calibrated by ALMA/NAASC staff following standard procedures. Each SPW was further self-calibrated in phase and amplitude in CASA 4.2.2. The self-calibrated phase and gain SPW-specific solutions were applied to create individual SPW line data cubes using the Briggs weighting scheme, with Brigg's robustness parameter set to 1–2. For the weak CH_3CN lines, a taper of 1″ was applied to the visibilities to maximize the signal-to-noise. The data cubes were subsequently continuum subtracted using channels identified as free of spectral line emission. The resulting r.m.s. noise in 1 km s^{-1} bins was ~2.3 mJy per beam. A CLEAN deconvolution was performed with a mask produced manually based on the $H^{13}CN$ emission and applied to the HC_3N and CH_3CN spectral line cubes (after verifying that the HC_3N and CH_3CN cubes contained no additional emission). For the 257 GHz continuum map, obtained by combining all SPWs in the USB, the synthesized beam is 0.65″ × 0.39″ (position angle PA = −7°) and r.m.s. noise 0.3 mJy per beam. The 257 GHz continuum flux density was determined to be 331 ± 33 mJy (assuming a 10% uncertainty in the absolute calibration), which agrees well with previous Submillimeter Array results[30].

Integrated emission maps were produced using the *immoments* task in CASA, adding emission $>1\sigma$ from each channel. Lower clipping levels do not change the result at the 10% level. Spectra were extracted using both the empirical CLEAN masks and theoretical masks based on Keplerian rotation to isolate disk emission in each channel. The CLEAN masks resulted in the best signal-to-noise ratio, and these results are used in the study.

Parametric disk density and temperature model. We adopt a previously developed disk density structure of MWC 480[10]. The total surface density is given by the similarity solution, which self-consistently describes the density structure of an accretion disk[31,32]:

$$\Sigma(r) = \Sigma_c \left(\frac{r}{R_c}\right)^{-\gamma} \exp\left[-\left(\frac{r}{R_c}\right)^{2-\gamma}\right] \tag{1}$$

where r is the radius (that is, the distance from the central star), $\Sigma_c = M_{gas}(2-\gamma)/(2\pi R_c^2)$ is a normalization coefficient, R_c is the characteristic radius, and γ is a gradient parameter describing the density fall-off with radius. The two-dimensional density structure in cylindrical coordinates is

$$\rho(r,z) = \frac{\Sigma_r}{H(r)\pi} \exp\left(\frac{-z^2}{H(r)^2}\right) \tag{2}$$

where $H(r) = H_{100}(r/100\ \mathrm{AU})^h$ is the scale height, H_{100} the height at a radius of 100 AU, and h a parameter that describes the amount of flaring of the disk. The density structure parameters were constrained using high angular resolution observations of the MWC 480 dust continuum emission at 1.3 mm and 3 mm wavelengths[10] and are listed in Extended Data Table 1. The resulting mass surface densities (in g cm^{-2}) are 550, 91 and 4.7 at radii r of 1, 10 and 100 AU, respectively.

Based on existing model fits to dust and ^{13}CO observations[33,34], the disk midplane temperature is parameterized as $T(r) = T_0\left(\frac{r}{R_0}\right)^{-q}$, where T_0 is set to 23 K at $R_0 = 100$ AU, and the power-law index q is 0.4. To account for the presence of a vertical temperature gradient due to heating of the disk surface by the central star[34,35], we parameterize the two-dimensional temperature structure in cylindrical coordinates as

$$T(r,z) = T_0\left(\frac{r}{R_0}\right)^{-q} \exp\left(\log\beta \frac{z}{H(r)}\right) \tag{3}$$

where T_0 is the midplane temperature at R_0, q is a power-law index describing the decrease in midplane temperature with radius, and β is a factor describing the increase in temperature at increasing disk height, which is set to 1.5 (ref. 34). The density and temperature disk structures are shown in Extended Data Fig. 1.

Molecular abundance retrieval. To retrieve molecular abundance and column density profiles, we define parametric abundance models with respect to the adopted MWC 480 disk density and temperature structure. We use a simple power-law prescription $x = x_{100AU} \times (r/100\ \mathrm{AU})^{\alpha}$ for the molecular abundances, where x is the abundance of the molecule with respect to the total hydrogen density, x_{100AU} is the abundance at 100 AU, r is the disk radius in AU, and α is a power-law index[36]. To obtain a column density profile, which is more intuitively related to observations, the abundance prescription is multiplied by disk surface density profile, equation (1). We set an outer cut-off radius (R_{out}) of 100 AU, corresponding to the most extended emission observed for any of the cyanides. We also explored R_{out} of 60 and 80 AU, but this did not improve the fit for any of the molecules. For each molecule, we calculate a grid of abundance models that covers x_{100AU} from 10^{-14} to 10^{-10} with respect to n_H, and α between 0 and 1 for $H^{13}CN$, 1 and 2 for HC_3N, 0 and 2 for CH_3CN. These initial ranges of α were selected on the basis of visual inspection of the observed emission, noting its level of central concentration, that is, $\alpha = 0$ corresponds to a steeply decreasing column density profile, $\alpha = 1$ to an almost flat profile, and $\alpha = 2$ to an increasing profile with radius.

The best-fit models are obtained by minimizing χ^2, the weighted difference between the real and imaginary part of the complex visibility measured in the (u, v)-plane sampled by the ALMA observations[36–38]. We use the three-dimensional Monte Carlo code LIME[39] to calculate the radiative transfer and molecular excitation. In addition to the molecular abundance profile and the disk density and temperature structure parameters, the radiative transfer modelling requires information on disk inclination, turbulence and velocity field. We adopted a disk inclination of 37°, disk turbulence of 0.05 km s^{-1} and a Keplerian velocity field with a stellar mass of 1.8 solar masses (ref. 10). For HCN and HC_3N, the level populations were computed in non-LTE (local thermodynamic equilibrium) with collision rates listed in the BASECOL database[40,41]. For CH_3CN, we assumed LTE, which is a reasonable approximation, since the expected disk emissive layers have higher densities than the critical density. We checked this assumption by running several HCN and HC_3N models assuming LTE and found a good agreement (within 20%) between LTE and non-LTE models. In Extended Data Fig. 2 we show the best-fit models for different α. In the case of HC_3N, only the models with $\alpha = 2$ show the observed ring-like structure and are therefore the only ones considered further.

Within the adopted model framework, the molecular abundances and abundance ratios are constrained within a factor of a few. These models implicitly assume a constant abundance profile with disk height, which may be a poor approximation in the outer disk where freeze-out in the midplane results in a vertically layered structure[20,42,43]. The existing data are, however, insufficient to put any constraints on the vertical distribution. Rather than pursuing a more complex model, we opted to test the sensitivity of the derived abundances and abundance ratios on the details of the vertical structure assumptions. We ran a second grid of models to simulate the effect of freeze-out in the midplane by reducing the molecular abundances by three orders of magnitude at $z/R < 0.2$, where z is the disk height and R the disk radius. Using the same fitting procedure as before, we find best-fit x_{100au} of 1×10^{-11} ($\alpha = 0$), 3.5×10^{-10} ($\alpha = 2$), and 5×10^{-11} ($\alpha = 1$), for $H^{13}CN$, HC_3N and CH_3CN, respectively. These abundances are an order of magnitude higher than those obtained assuming a vertically constant abundance. The $CH_3CN/H^{13}CN$ abundance ratio profile is unchanged within the model uncertainties, however. Despite a higher excitation level of CH_3CN the two molecules emit from a similar disk layer when assigned the same vertical abundance profile due to a combination of a high critical density of the CH_3CN transition and a rapid fall-off in density with disk height, that is, practically no CH_3CN emission originates at $z/R > 0.5$. In contrast, the derived HC_3N/HCN ratio does depend on the vertical structure. The HC_3N transition has an even higher energy level and a lower critical density than CH_3CN, which implies that some of the HC_3N emission can originate in more elevated disk layers than $H^{13}CN$ and CH_3CN. The reported $CH_3CN/H^{13}CN$ results are thus robust to the abundance model assumptions, while the reported $HC_3N/H^{13}CN$ results may be overestimated by up to an order magnitude.

Disk chemistry modelling to constrain the ice-to-gas ratios. To constrain the origin of CH_3CN and therefore how its ice-to-gas ratio compares with HCN, we first explore how much of the observed CH_3CN gas could come from gas-phase chemistry. We ran a grid of pseudo-time-dependent models with a complete gas-phase chemistry[44], but without any grain surface reactions (except for adsorption and desorption, and H_2 formation on grains), for 1 Myr. The grid covers several orders of magnitude of densities, temperatures, ultraviolet fields and ionization fractions—the four most important regulators of gas-phase chemistry in disks. For gas-phase CH_3CN production, the ionization rate is expected to be of special importance because the main gas-phase formation pathway includes CH_3^+ (through $HCN + CH_3^+$, which has a expert-validated reaction rate in the KIDA database[45]). Extended Data Fig. 3 shows that none of these models can produce CH_3CN/HCN > 0.01 and that in the vast majority of the parameter space CH_3CN/HCN < 0.001. A ratio of 0.01 is approached for an ionization rate of 10^{-14} s^{-1}, but such high ionization rates are only attained in the disk atmosphere where stellar X-rays regulate the ionization balance. This disk layer contain a very small fraction of the total disk mass and therefore contributes a negligible amount of the total molecular column. Deeper into the disk, the ionization rate is reduced below 10^{-17} s^{-1} because

of attenuation of both X-rays and cosmic rays[46]. A ratio of 0.01 is also approached at lower ionization rates at very low temperatures and high densities. Such environments are characteristic for the outer disk midplane, which contains a lot of mass, but freeze-out at these temperatures results in CH_3CN abundances below 10^{-14}, which would not contribute to the observed CH_3CN emission. The observed high abundance of CH_3CN gas with respect to HCN therefore implies that grain surface chemistry contributes significantly to the observed CH_3CN abundance.

We ran a number of complete disk chemistry models for 1 Myr, the estimated age of the MWC 480 star+disk system, which include gas and grain surface chemistry, as well as different levels of turbulence, to calculate the range of plausible ice-to-gas conversion factors for HCN (and other gas-phase chemistry products) and CH_3CN (ref. 47). The main grain surface CH_3CN formation pathways in these models are hydrogenation of C_2N and CH_3+CN and ice photochemistry. The main desorption pathway is ultraviolet photodesorption because of the high binding energies of HCN, HC_3N and CH_3CN: 4,170, 4,580 and 4,680 K, respectively[48–50], corresponding to sublimation temperatures of ~90 K at a gas density of 10^6 cm^{-3} and 110 K at a gas density of 10^{10} cm^{-3}. The three cyanides should also present very similar ultraviolet photodesorption efficiencies because of similar radiation cross-sections and binding energies[51]. Without grain surface chemistry, these molecules would thus have had comparable ice-to-gas ratios.

Because CH_3CN can form on the grains in the cold midplane, a key feature is the coupling between the ice formation layer and the disk layers where molecules are efficiently desorbed into the gas phase. Without such a coupling, ice chemistry cannot affect the gas-phase volatile composition. In the disk models, we define the vertical diffusion coefficient as $D_z = \alpha_z c_s^2/\Omega$, where α_z is a free parameter, c_s is the local sound velocity, and Ω is the Keplerian orbital frequency. Different levels of turbulence are simulated by different values of α_z between 0 and 0.01. In models with turbulence turned off ($\alpha_z = 0$), there is no mixing between the cold, icy midplane and the intermediate layers where gas-phase molecules are abundant (and the observed cyanides probably reside). Extended Data Fig. 4a shows that without turbulence the predicted CH_3CN abundances are low everywhere ($<10^{-11}$ with respect to n_H, where n_H is the number density of H nuclei) and the vertically averaged CH_3CN/HCN abundance ratio is <0.001; see Extended Data Fig. 5a, d.

In models that include vertical mixing ($\alpha_z = 0.001$–0.01), the CH_3CN abundance as well as the CH_3CN/HCN abundance ratio are enhanced by an order of magnitude because of mixing of midplane icy grains into ultraviolet exposed disk regions where the ices photodesorb (Extended Data Figs 4 and 5). In particular, the CH_3CN/HCN ratio approaches 10^{-2} outside of 30 AU, within an order of magnitude of the observed ratios. Model–data comparison thus suggest that the intermediate disk layers are strongly coupled to the icy midplane through mixing. This implies that gas-phase observations of CH_3CN and other cyanides at intermediate disk heights can be used to trace the total volatile reservoir, even if the conversion from gas to ice can be complex.

Using the three models, with no, moderate and strong mixing, we calculate the HCN and CH_3CN ice-to-gas density and column density ratios as a function of disk radius (Extended Data Figs 6 and 7). For HCN, the ice-to-gas ratio varies between 10^5 and 10^3 between 30 and 100 AU. For CH_3CN, the ice-to-gas ratio varies between 10^7 and 10^4. In other disk chemistry models[18,52] the contrast between the two ice-to-gas ratios is even higher. Considering the range of values and their sensitive dependence on assumptions about disk turbulence, we conclude that the ice-to-gas ratio for CH_3CN is at least 10 times higher than HCN and HC_3N. At longer timescales the balance between formation in the midplane and destruction at more exposed disk layers determines whether the CH_3CN abundance increases or decreases—in general, the more turbulence the faster the destruction.

It is important to note that we employ generic disk structures[53] that have not been fitted to the MWC 480 data. The temperature and density structures are comparable (compare Extended Data Figs 1 and 4). The ultraviolet field is lower than would be expected towards a Herbig Ae star and the gas-phase CH_3CN production may therefore be somewhat overestimated in our model, but as shown above, gas-phase chemistry is a minor contributor to the overall CH_3CN gas budget. The grain surface production takes place in ultraviolet shielded regions and should thus not be affected by an increased ultraviolet flux. An increased ultraviolet field intensity may change the desorption/adsorption balance, however, since ultraviolet photodesorption is an important desorption mechanism. This increase in photodesorption will be accompanied by an increased level of photodissociation and the main effect will therefore be to push the desorbing layer closer to the midplane rather than increasing its abundance. An A star type radiation field may aid the detection of gas-phase CH_3CN, since less turbulence may be required to bring the icy grains up to ultraviolet exposure and into the gas phase. A decrease in the disk height of the cyanide gas layer, compared to our generic disk model, may also reduce the absolute ice-to-gas ratios, but the relative ratios of species with similar adsorption and desorption characteristics should not be significantly affected.

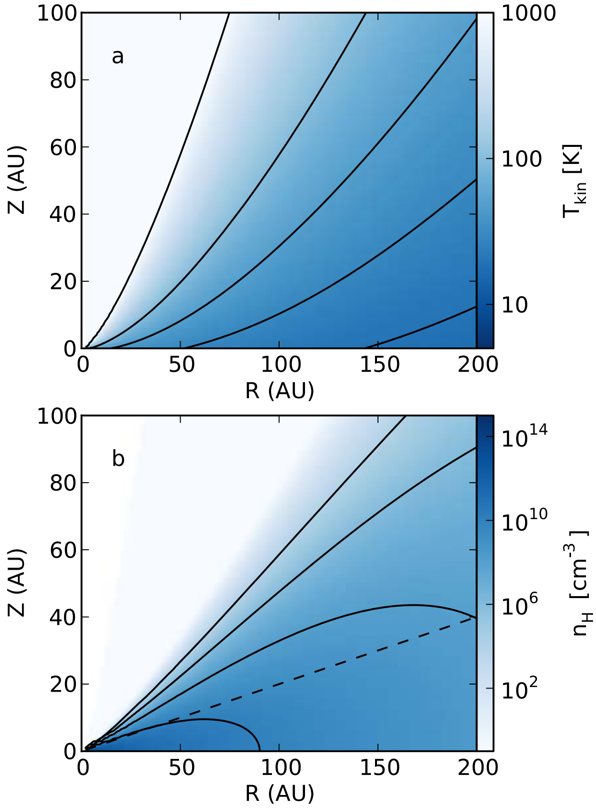
Desorption mechanism. In our disk chemistry model, ultraviolet photodesorption regulates the desorption of HCN, HC_3N and CH_3CN ice into the gas phase. This result depends on a combination of disk and star properties and on the theoretical yields of different desorption mechanisms, many of which are poorly constrained. While relative ice-to-gas ratios of related species are insensitive to the details of the desorption process, the absolute ice-to-gas ratios found in our models must be considered highly uncertain. Based on theory, different desorption mechanisms should present different radial and vertical abundance distributions, which could be tested observationally. Desorption due to release of chemical formation energy should be most important in the cold disk midplane, while ultraviolet photodesorption is only efficient in warmer, upper disk layers, and thermal desorption in the disk layers that are warm enough. In particular, colder molecular emission layers should be characterized by lower excitation temperatures, which can be measured when multiple lines of the same species are observed. Observational constraints on ice desorption in disks will thus be attainable with ALMA.

Different desorption mechanisms should also exhibit different chemical signatures, although the details are often model-dependent. In the case of thermal desorption, the gas phase composition in a layer should simply depend on volatility, albeit with the added complication of ice entrapment[54]. Chemistry can also help distinguish between ice and gas origins and thus provide an independent measure of the relative importance of the two pathways. The high ratio of CH_3CN/CH_3NC in the Horsehead PDR (Photon Dominated Region), for example, was used as evidence for a grain surface origin of the observed CH_3CN (ref. 55). Chemical tracers could also provide useful constraints on the incident radiation fields and temperature structure and thus aid in ruling out some of the potentially important desorption mechanisms. For example, CN/HCN is a proposed tracer of ultraviolet flux and HNC/HCN of temperature[56,57].

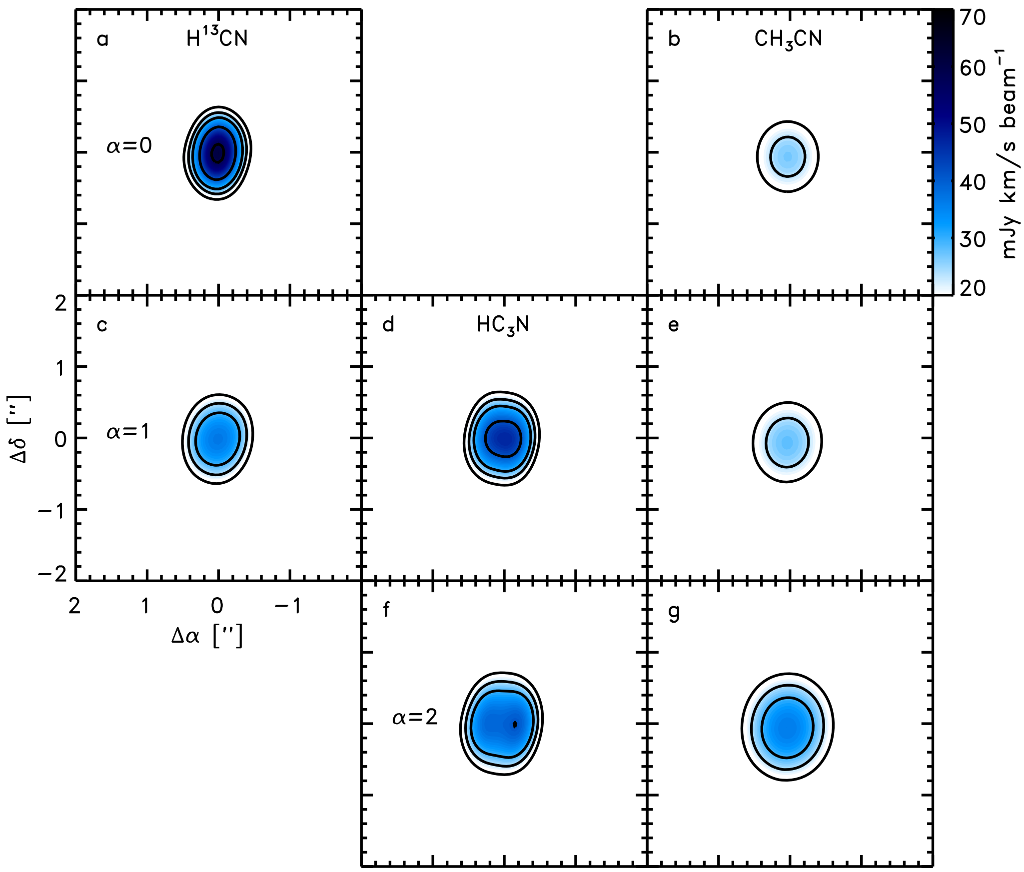
Sample size. No statistical methods were used to predetermine sample size.

30. Öberg, K. I. *et al.* Disk imaging survey of chemistry with SMA. II. Southern sky protoplanetary disk data and full sample statistics. *Astrophys. J.* **734,** 98 (2011).
31. Lynden-Bell, D. & Pringle, J. E. The evolution of viscous discs and the origin of the nebular variables. *Mon. Not. R. Astron. Soc.* **168,** 603–637 (1974).
32. Hartmann, L., Calvet, N., Gullbring, E. & D'Alessio, P. Accretion and the evolution of T Tauri disks. *Astrophys. J.* **495,** 385–400 (1998).
33. Piétu, V., Dutrey, A., Guilloteau, S., Chapillon, E. & Pety, J. Resolving the inner dust disks surrounding LkCa 15 and MWC 480 at mm wavelengths. *Astron. Astrophys.* **460,** L43–L47 (2006).
34. Dartois, E., Dutrey, A. & Guilloteau, S. Structure of the DM Tau outer disk: probing the vertical kinetic temperature gradient. *Astron. Astrophys.* **399,** 773–787 (2003).
35. Rosenfeld, K. A., Andrews, S. M., Wilner, D. J., Kastner, J. H. & McClure, M. K. The structure of the evolved circumbinary disk around V4046 Sgr. *Astrophys. J.* **775,** 136 (2013).
36. Qi, C. *et al.* Imaging of the CO snow line in a solar nebula analog. *Science* **341,** 630–632 (2013).
37. Qi, C., Wilner, D. J., Aikawa, Y., Blake, G. A. & Hogerheijde, M. R. Resolving the chemistry in the disk of TW Hydrae. I. Deuterated species. *Astrophys. J.* **681,** 1396–1407 (2008).
38. Qi, C. *et al.* Resolving the CO snow line in the disk around HD 163296. *Astrophys. J.* **740,** 84 (2011).
39. Brinch, C. & Hogerheijde, M. R. LIME — a flexible, non-LTE line excitation and radiation transfer method for millimeter and far-infrared wavelengths. *Astron. Astrophys.* **523,** A25 (2010).
40. Wernli, M., Wiesenfeld, L., Faure, A. & Valiron, P. Rotational excitation of HC_3N by H_2 and He at low temperatures. *Astron. Astrophys.* **464,** 1147–1154 (2007).
41. Dubernet, M., Nenadovic, L. & Doronin, N. in *Astronomical Data Analysis Software and Systems XXI* (eds Ballester, P., Egret, D. & Lorente, N. P. F.) 335–338 (Conf. Ser. Vol. 461, Astronomical Society of the Pacific, 2012).
42. Aikawa, Y. & Herbst, E. Molecular evolution in protoplanetary disks. Two-dimensional distributions and column densities of gaseous molecules. *Astron. Astrophys.* **351,** 233–246 (1999).
43. Walsh, C., Millar, T. J. & Nomura, H. Molecular line emission from a protoplanetary disk irradiated externally by a nearby massive star. *Astrophys. J.* **766,** L23 (2013).
44. Aikawa, Y., Wakelam, V., Hersant, F., Garrod, R. T. & Herbst, E. From prestellar to protostellar cores. II. Time dependence and deuterium fractionation. *Astrophys. J.* **760,** 40 (2012).
45. Wakelam, V. *et al.* A kinetic database for astrochemistry (KIDA). *Astrophys. J. Suppl. Ser.* **199,** 21 (2012).
46. Cleeves, L. I., Bergin, E. A., Qi, C., Adams, F. C. & Oberg, K. I. Constraining the X-ray and cosmic ray ionization chemistry of the TW Hya protoplanetary disk: evidence for a sub-interstellar cosmic ray rate. *Astrophys. J.* **799,** 204 (2015).
47. Furuya, K. & Aikawa, Y. Reprocessing of ices in turbulent protoplanetary disks: carbon and nitrogen chemistry. *Astrophys. J.* **790,** 97 (2014).
48. Yamamoto, T., Nakagawa, N. & Fukui, Y. The chemical composition and thermal history of the ice of a cometary nucleus. *Astron. Astrophys.* **122,** 171–176 (1983).
49. Garrod, R. T. & Herbst, E. Formation of methyl formate and other organic species in the warm-up phase of hot molecular cores. *Astron. Astrophys.* **457,** 927–936 (2006).
50. Collings, M. P. *et al.* A laboratory survey of the thermal desorption of astrophysically relevant molecules. *Mon. Not. R. Astron. Soc.* **354,** 1133–1140 (2004).

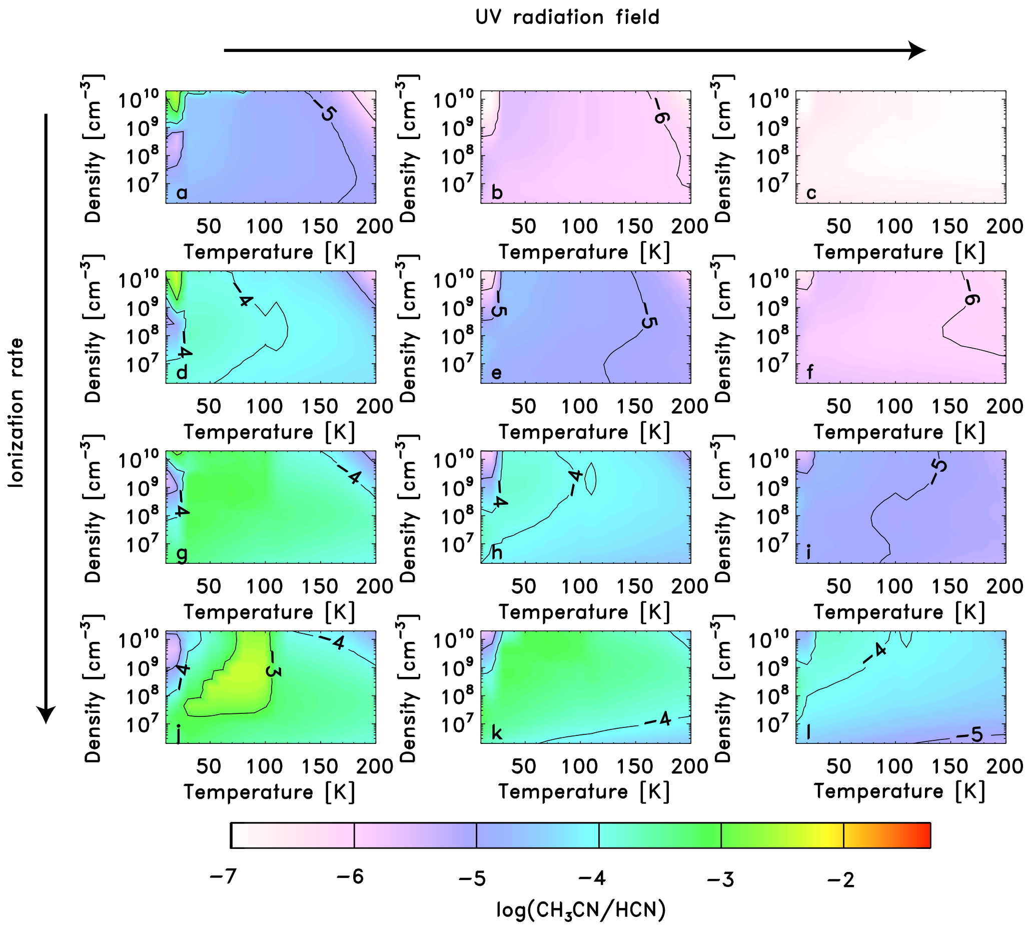
51. van Dishoeck, E. F., Jonkheid, B. & van Hemert, M. C. Photoprocesses in protoplanetary disks. *Faraday Discuss.* **133,** 231–243 (2006).
52. Walsh, C., Nomura, H., Millar, T. J. & Aikawa, Y. Chemical processes in protoplanetary disks. II. On the importance of photochemistry and X-ray ionization. *Astrophys. J.* **747,** 114 (2012).
53. Nomura, H., Aikawa, Y., Tsujimoto, M., Nakagawa, Y. & Millar, T. J. Molecular hydrogen emission from protoplanetary disks. II. Effects of X-ray irradiation and dust evolution. *Astrophys. J.* **661,** 334–353 (2007).
54. Fayolle, E. C., Öberg, K. I., Cuppen, H. M., Visser, R. & Linnartz, H. Laboratory H_2O:CO_2 ice desorption data: entrapment dependencies and its parameterization with an extended three-phase model. *Astron. Astrophys.* **529,** A74 (2011).
55. Gratier, P. *et al.* The IRAM-30 m line survey of the Horsehead PDR. III. High abundance of complex (iso-)nitrile molecules in UV-illuminated gas. *Astron. Astrophys.* **557,** A101 (2013).
56. Bergin, E., Calvet, N., D'Alessio, P. & Herczeg, G. J. The effects of UV continuum and Lyα radiation on the chemical equilibrium of T Tauri disks. *Astrophys. J.* **591,** L159–L162 (2003).
57. Graninger, D. M., Herbst, E., Öberg, K. I. & Vasyunin, A. I. The HNC/HCN ratio in star-forming regions. *Astrophys. J.* **787,** 74 (2014).



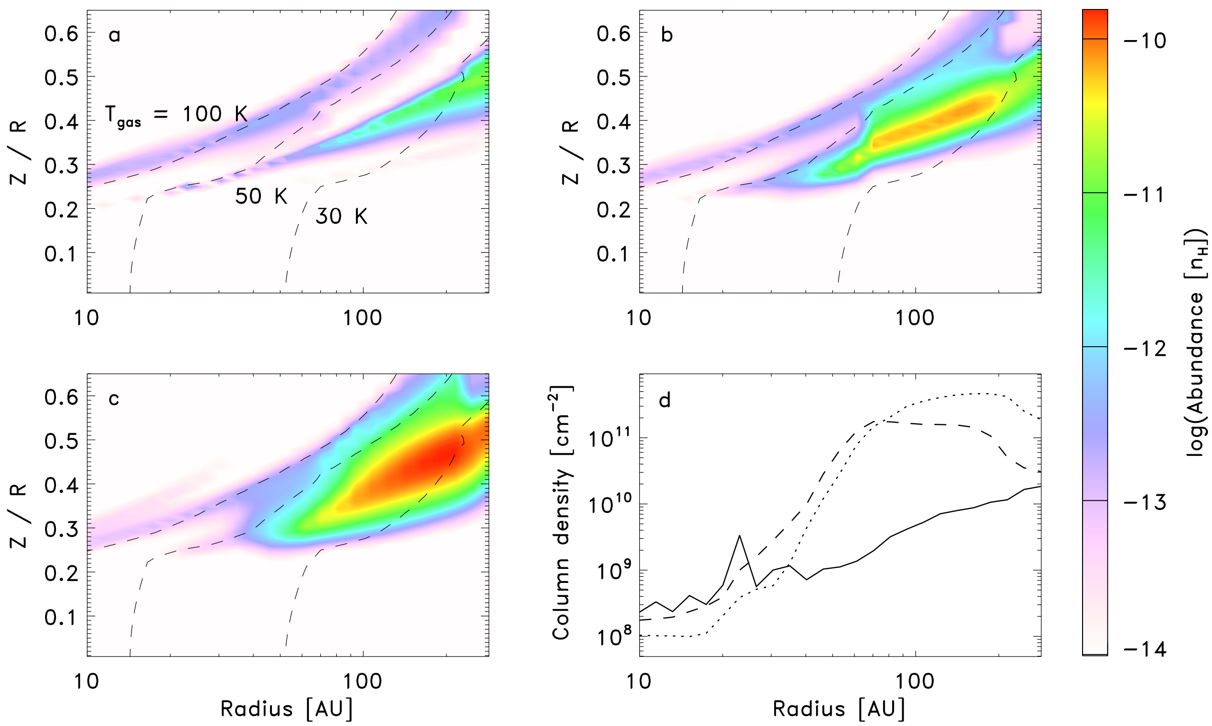
Extended Data Figure 1 | Model of the physical structure of the MWC 480 protoplanetary disk. **a**, Radial (distance R) and vertical (distance Z) disk temperature profile (colour: see colour scale on right, contours: the gas temperature $T_{\mathrm{kin}} = 20$, 30, 50, 100 and 1,000 K). **b**, Radial (R) and vertical (Z) density profile (colour: see colour scale on the right, contours: hydrogen density n_{H} 10^{10}, 10^8, 10^6 and 10^4 cm^{-3}). $Z/R = 0.2$ is marked with a dashed line.



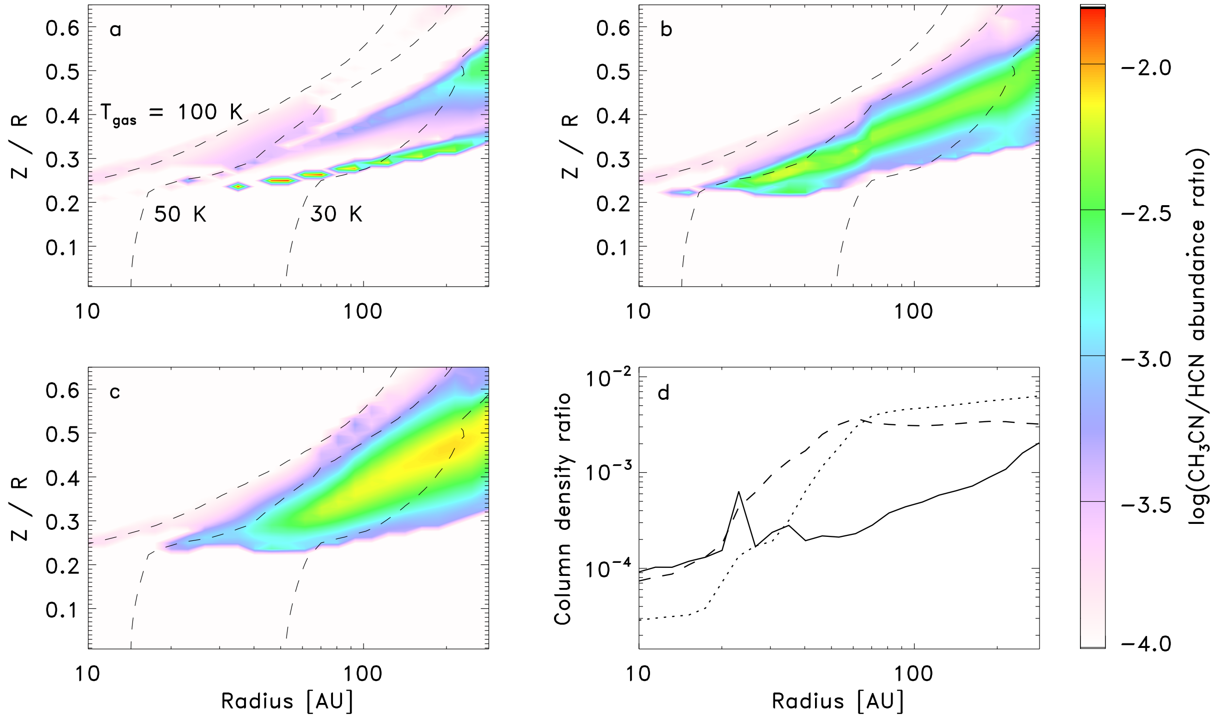
Extended Data Figure 2 | Synthetic observations of $H^{13}CN$, HC_3N and CH_3CN for different density slopes α. The models are based on best fit to data for different choices of α, with the ranges chosen based on the emission pattern for each molecule. Left column, $H^{13}CN$; middle column, HC_3N; right column, CH_3CN. Top row, $\alpha = 0$; middle row, $\alpha = 1$; bottom row, $\alpha = 2$. **a–g**, Integrated emission maps (colour: see colour scale on the right). Black contours are the observed $[3, 4, 5, 7, 10]\sigma$ in Fig. 1. The synthesized beam is shown in the bottom left corner of each panel. Note the change in emission profile between $\alpha = 1$ and 2 for HC_3N.



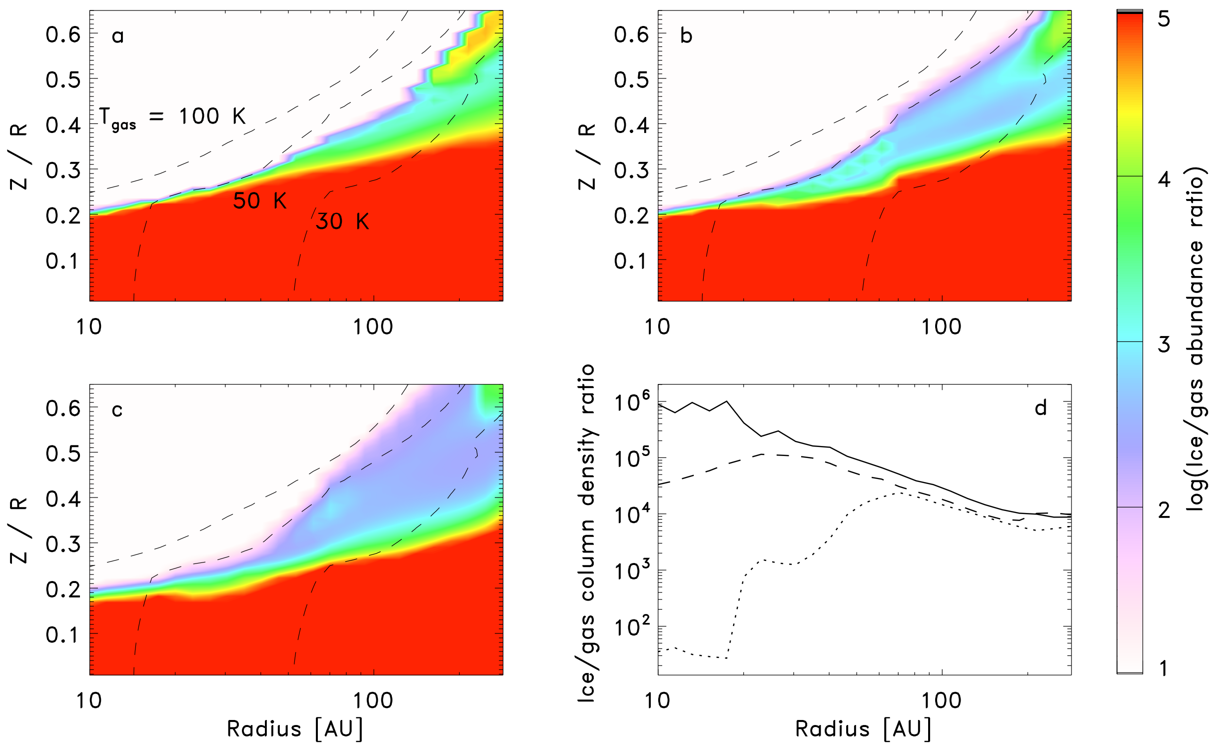
Extended Data Figure 3 | Models of gaseous CH_3CN/HCN abundance ratios under different physical conditions. **a**–**l**, The CH_3CN/HCN abundance ratio on a logarithmic scale (colour: see colour scale on the bottom and numbers on contours). The ultraviolet radiation flux increases from left to right from $G_0 = 1$ (**a**, **d**, **g**, **j**) to $G_0 = 10$ (**b**, **e**, **h**, **k**) to $G_0 = 100$ (**c**, **f**, **i**, **l**), where G_0 is the scaling factor in multiples of the local interstellar radiation field. The ionization rate of H_2 increases from top to bottom from $10^{-17}\,s^{-1}$ (**a**–**c**) to $10^{-16}\,s^{-1}$ (**d**–**f**) to $10^{-15}\,s^{-1}$ (**g**–**i**) to $10^{-14}\,s^{-1}$ (**j**–**l**).



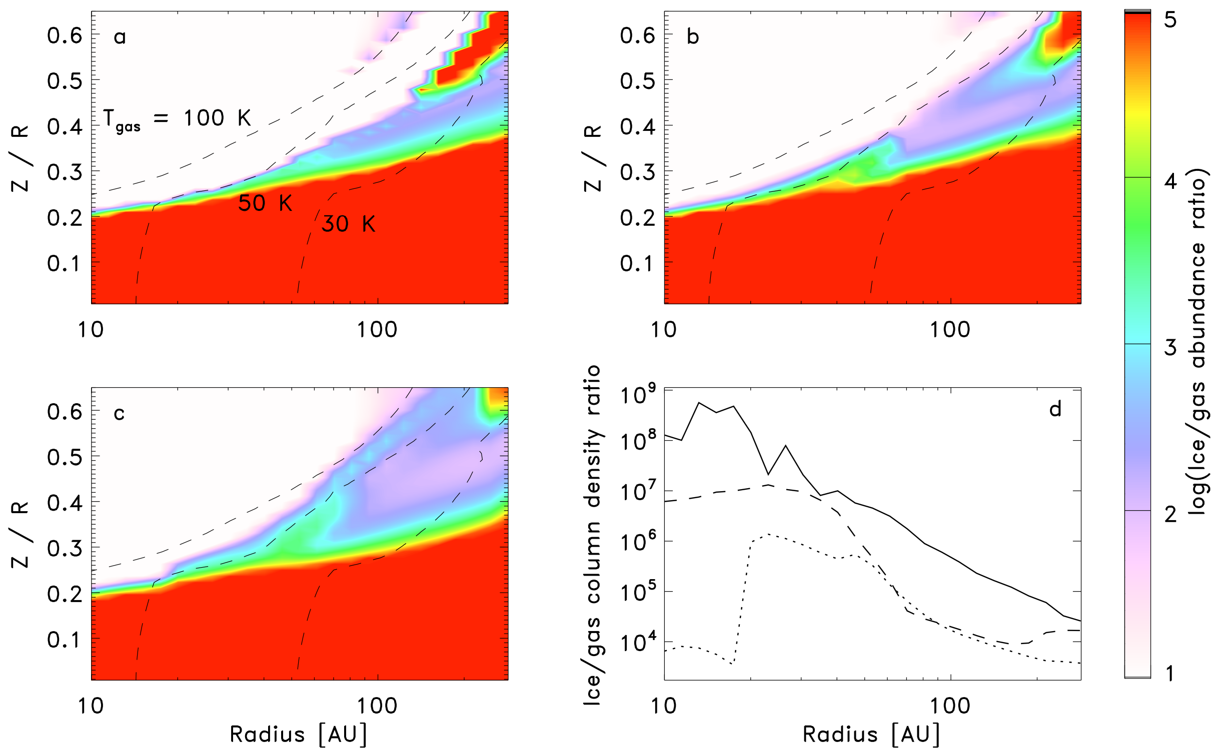
Extended Data Figure 4 | Models of gaseous CH_3CN in disks with and without turbulent diffusion. **a**, The abundance of CH_3CN with respect to the hydrogen density n_H (colour: see colour scale on the right) as a function of disk radius (R) and height scaled by the radius (Z/R) in a model without turbulence. The dashed lines indicate gas temperatures of [30, 50, 100] K. **b**, **c**, As **a** but in disk models that include turbulence parameterized by $\alpha_z = 10^{-3}$ (**b**) and $\alpha_z = 10^{-2}$ (**c**). **d**, The vertically integrated column density of CH_3CN from **a**–**c** (solid line: $\alpha_z = 0$, dashed line: $\alpha_z = 10^{-3}$, dotted line: $\alpha_z = 10^{-2}$).



Extended Data Figure 5 | Models of gaseous CH_3CN/HCN ratios in disks with and without turbulent diffusion. a–d, As in Extended Data Fig. 4 but for CH_3CN/HCN ratio.



Extended Data Figure 6 | Models of gas-to-ice ratios of HCN in disks with and without turbulent diffusion. a–d, As in Extended Data Fig. 4 but for ice-to-gas ratios of HCN.



Extended Data Figure 7 | Models of gas-to-ice ratios of CH_3CN in disks with and without turbulent diffusion. a–d, As in Extended Data Fig. 4 but for ice-to-gas ratios of CH_3CN.

Extended Data Table 1 | Physical model for the disk of MWC 480

Parameters	Values
Stellar properties	
Estimated distance: d (pc)	140
Stellar mass: M_* ($M_\odot$)	1.8
Disk structure properties	
Disk mass: M_d($M_\odot$)	0.18
Characteristic radius: R_c(AU)	81
Outer cut-off radius (AU)	100
Scale height: $H_{100\mathrm{AU}}$ (AU)	16
Flaring index: h	1.25
Density power-law index: γ	0.75
Midplane temperature: $T_{100\mathrm{AU}}$ (K)	23
Temperature power-law index: q	0.5
Vertical temperature gradient index: β	1.5
Disk geometric and kinematic properties	
Inclination: i (deg)	37
Systemic velocity: V_{LSR}(km s^{-1})	5.0
Turbulent line width: v_{turb}(km s^{-1})	0.05
Position angle: P.A.(deg)	58

doi:10.1038/nature14278

Saturn's fast spin determined from its gravitational field and oblateness

Ravit Helled[1], Eli Galanti[2] & Yohai Kaspi[2]

The alignment of Saturn's magnetic pole with its rotation axis precludes the use of magnetic field measurements to determine its rotation period[1]. The period was previously determined from radio measurements by the Voyager spacecraft to be 10 h 39 min 22.4 s (ref. 2). When the Cassini spacecraft measured a period of 10 h 47 min 6 s, which was additionally found to change between sequential measurements[3,4,5], it became clear that the radio period could not be used to determine the bulk planetary rotation period. Estimates based upon Saturn's measured wind fields have increased the uncertainty even more, giving numbers smaller than the Voyager rotation period, and at present Saturn's rotation period is thought to be between 10 h 32 min and 10 h 47 min, which is unsatisfactory for such a fundamental property. Here we report a period of 10 h 32 min 45 s ± 46 s, based upon an optimization approach using Saturn's measured gravitational field and limits on the observed shape and possible internal density profiles. Moreover, even when solely using the constraints from its gravitational field, the rotation period can be inferred with a precision of several minutes. To validate our method, we applied the same procedure to Jupiter and correctly recovered its well-known rotation period.

Previous theoretical attempts to infer Saturn's rotation period have relied on wind observations derived from cloud tracking at the observed cloud level[6]. One theoretical approach was based on minimizing the 100 mbar dynamical heights[7] with respect to Saturn's measured shape[8], while a second approach was based on analysing the potential vorticity in Saturn's atmosphere from its measured wind profile[9]. The derived rotation periods were found to be 10 h 32 min 35 s ± 13 s, and 10 h 34 min 13 s ± 20 s, respectively. Our optimization method is based on linking the rotation period of Saturn with its observed physical properties and their uncertainties, in particular, the gravitational field. The method allows us to derive Saturn's rotation period for different types of constraints, and does not rely on a specific interior model, equation of state, wind properties, or other indirect measurements.

The gravitational moments and the internal density profile can be related through the smallness parameter $m = \omega^2 R^3/GM$, where R is the planet's mean radius, M is its mass, G is the gravitational constant, and $\omega = 2\pi/P$ is the angular velocity associated with the rotation period P (refs 10 and 11). The even gravitational moments can be expanded as a function of m by $J_{2n} = \sum_{k=n}^{3} m^k a_{2n,k}$, where $a_{2n,k}$ are coefficients that are determined by the radial density distribution (see Methods). The expansion can go to any order of n; since at present only J_2, J_4 and J_6 are known for Saturn (and Jupiter), in this study we take $n = 1, 2, 3$.

The relation for J_{2n} shows that the measured gravitational moments are determined from the combination of the internal density distribution ($a_{2n,k}$) as well as rotation (m). Our goal is to find a solution for m and $a_{2n,k}$ that minimizes the difference between the observed and calculated J_{2n} within the observed uncertainties. The $a_{2n,k}$ can be expressed by a combination of figure functions (see Methods) that represent a given internal density profile, and can then be linked to the gravitational moments[10,12,13]. However, since in this case there are only three equations and seven unknowns (six figure functions and the smallness parameter), there is no unique solution. As a result, the solution is found by using a statistical optimization approach.

We define an optimization function as the sum of the normalized absolute differences between the observed gravitational moments and the calculated gravitational moments, given by:

$$Y = \sum \left(\frac{\left|J_2 - J_2^{\text{obs}}\right|}{\left|\Delta J_2^{\text{obs}}\right|} + \frac{\left|J_4 - J_4^{\text{obs}}\right|}{\left|\Delta J_4^{\text{obs}}\right|} + \frac{\left|J_6 - J_6^{\text{obs}}\right|}{\left|\Delta J_6^{\text{obs}}\right|} \right) \tag{1}$$

where J_{2n} are the calculated moments, J_{2n}^{obs} are the measured moments, and $\Delta J_{2n}^{\text{obs}}$ are the measurement uncertainties of the measured gravitational moments[14,15]. The optimization procedure begins with an initial guess of the various parameters being randomly spread throughout the physical bounds of each parameter. This is repeated 2,000 times to achieve statistical significance. From these 2,000 cases we compute the rotation period and its standard deviation (see Methods). An example of the derived solutions using our optimization method is presented in Fig. 1.

The entire set of solutions for Saturn are summarized in Fig. 2 which shows P_{calc} (dots) and its 1σ standard deviation (blue shading). We first present solutions that are completely unconstrained in radius and density structure, and where the rotation period (grey shading) is allowed to vary widely (Fig. 2a). The fact that the calculated standard deviation is much smaller than the allowed range (blue shading being much narrower than the grey shading) indicates that knowledge of the gravitational

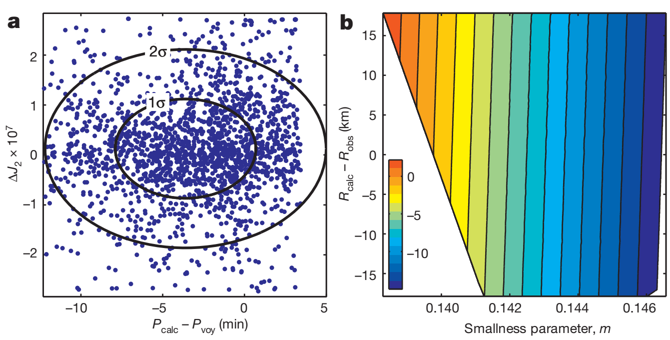


Figure 1 | An example of the statistical distribution of solutions for Saturn's rotation period. For this specific case, the initial possible range of rotation periods is taken to have an uncertainty of 0.5 h around the Voyager radio period (hereafter, P_{voy}). The calculated mean radius was set to be within 20 km of Saturn's observed mean radius. The solution is based on an ensemble of 2,000 individual sub-cases, each of them representing a case with specific random initial conditions within the defined parameter space. **a**, A scatter plot of the distribution of solutions on the plane of the calculated rotation period P_{calc} minus $P_{\text{voy}} = 10$ h 39 min 22 s and ΔJ_2. Each blue dot represents one sub-case solution. The inner and outer black circles show the first and second standard deviations, respectively. **b**, The distribution of the derived rotation period with respect to P_{voy} (in minutes) as a function of smallness parameter m and the calculated mean radius R_{calc} minus the observed mean radius of Saturn ($R_{\text{obs}} = 58,232$ km; ref. 7).

[1]Department of Geosciences, Raymond & Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, 69978, Israel. [2]Department of Earth and Planetary Sciences, Weizmann Institute of Science, Rehovot, 76100, Israel.

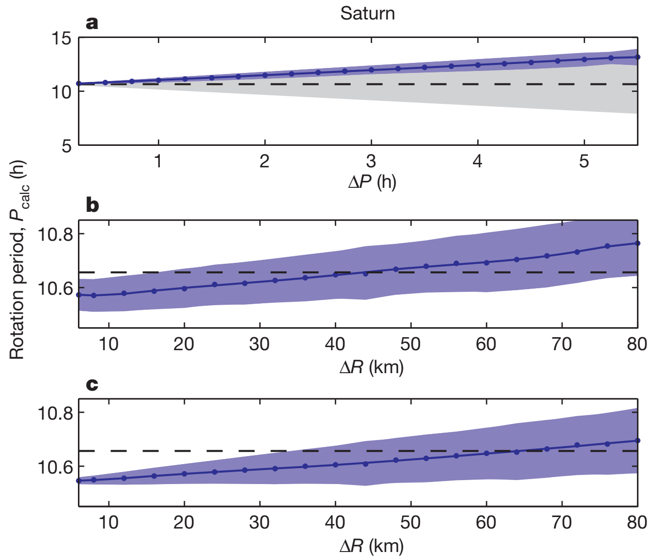


Figure 2 | Solutions for Saturn's rotation period. **a**, The calculated period P_{calc} (blue dots) and its 1σ standard deviation (blue shading) for a large range of cases for which the assumed possible range in rotation period varies between 0.25 h and 5.5 h (grey shading) around P_{voy} (black-dashed line). **b**, P_{calc} and its 1σ standard deviation (blue shading) using $\Delta P = 0.5$ h versus the assumed uncertainty in Saturn's observed mean radius R_{obs}. **c**, As for **b** but when the figure functions are also constrained.

moments can be used to narrow the possible range of rotation periods. In addition, as the initial range of the possible rotation periods is narrower, the derived rotation period can be determined with higher precision. For the smallest range in rotation period (left dot in Fig. 2a) we derive a rotation period of 10 h 43 min 10 s ± 4 min. The fact that the uncertainty in rotation period is decreased without enforcing tight constraints on the model emphasizes the strength of this method. Nonetheless, without any constraints on the shape the solution for the rotation period still has a relatively large range of solutions. In reality, occultation measurements[7,16,17] provide bounds on the shape of the planet (radius versus latitude), and as shown below this allows the rotation period to be further constrained[8,18,19].

The best measurement uncertainty of Saturn's radii from radio and stellar occultation is ~6 km (ref. 17), although the actual uncertainty could be larger owing to the unknown contribution of the atmospheric dynamics[6] to the measured shape[20]. We therefore explore a range of uncertainty in mean radius between 6 km and 80 km. The results for this case are shown in Fig. 2b where P_{calc} and its standard deviation versus the uncertainty in observed radius R_{obs} are shown. The standard deviation (blue shading) of P_{calc} decreases with decreasing uncertainty in the radius. For an uncertainty of 6 km in Saturn's mean radius we derive a rotation period of 10 h 34 min 22 s ± 3.5 min. It is important to note that the derived period when using only the gravitational field is larger than P_{voy} while the derived period with the shape constraint is faster than P_{voy} (see Methods and Fig. 3 in Extended Data). It is clear that the parameter space of possible solutions narrows when the constraint of Saturn's measured mean radius is included. Yet, geopotential variations caused by atmospheric dynamics affect the shape of the planet, and therefore care should be taken when considering these measurements. By taking this hierarchal approach we are able to isolate the uncertainty given estimates of shape and internal structure separately. More conservative uncertainties in radius (tens of kilometres) yield longer rotation periods, thus giving solutions closer to the Voyager rotation period (see Fig. 2b).

The uncertainty in P_{calc} can be decreased even further if we also limit the range of the figure functions, that is, the density profile (Fig. 2c). Limiting the figure functions to within a range implied by interior structure models[21,22] (see Methods), the derived period is found to be 10 h 32 min 45 s ± 46 s. This rotation period is in agreement with previous calculations that derived Saturn's rotation period by using a fit to Saturn's measured shape[8]. The fact that the rotation period is shorter than the Voyager rotation period also implies that the latitudinal wind structure is more symmetrical, thus containing both easterly and westerly jets as on Jupiter[9,23]. Although the smallest possible uncertainty in rotation period is desirable, there is a clear advantage in not specifying constraints on the density profile, and keeping the method as general as possible.

Unlike Saturn's, the rotation period of Jupiter is well determined owing to its tilted magnetic field. Jupiter's measured rotation period (system III) is 9 h 55 min 29.69 s (refs 24, 25). To verify the robustness of our results we apply this method also for Jupiter (Fig. 3). When only the gravitational moments are used as constraints (Fig. 3a), as for Saturn, the uncertainty in the calculated rotation period is much smaller than the allowed range, and converges towards Jupiter's rotation period. Figure 3b shows the sensitivity of the derived period, when the uncertainty in period is ±0.5 h around the measured value, for a range of possible mean radii. As for Saturn, the standard deviation of P_{calc} decreases with decreasing ΔR. When the variation in R_{obs} is taken to be 6 km, a rotation period of 9 h 56 min 6 s ± 1.5 min is derived, consistent with Jupiter's measured rotation period. When we also add constraints on the figure functions, the derived rotation period becomes 9 h 55 min 57 s ± 40 s, showing that our method reproduces Jupiter's rotation period successfully.

The determination of Saturn's J_{2n} is expected to improve substantially following Cassini's end-of-mission proximal orbits. To test whether a more accurate determination of Saturn's gravitational field will allow us to better constrain its rotation period, we repeat the optimization with the expected new uncertainty on the gravitational moments ($\Delta J_{2n} \approx 10^{-9}$)[26] around the currently measured values. The solution making no assumptions on the density profile is shown in Fig. 4a. Since Jupiter's gravitational field will be more tightly determined by the Juno spacecraft[26,27] we do a similar analysis for Jupiter (Fig. 4b). While for Jupiter the calculated rotation period remains the same with the more accurate gravitational field, for Saturn the calculated uncertainty of

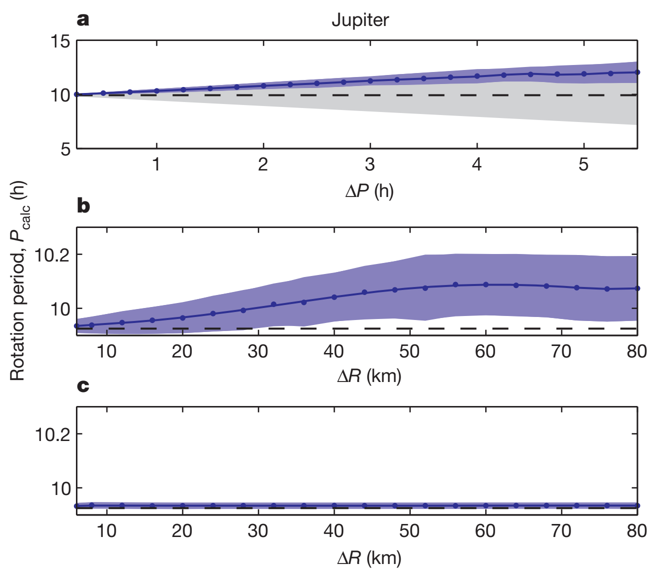


Figure 3 | Solutions for Jupiter's rotation period. **a**, The calculated period P_{calc} (blue dots) and its 1σ standard deviation (blue shading) for a large range of cases for which the assumed possible range in rotation period varies between 0.25 h and 5.5 h, that is, between ~5 h and 15 h (grey shading) around Jupiter's measured period (black-dashed line). **b**, P_{calc} and its 1σ standard deviation (blue shading) using $\Delta P = 0.5$ h versus the uncertainty in the assumed uncertainty in Jupiter's observed mean radius R_{obs}. **c**, As for **b** but when the figure functions are also constrained.

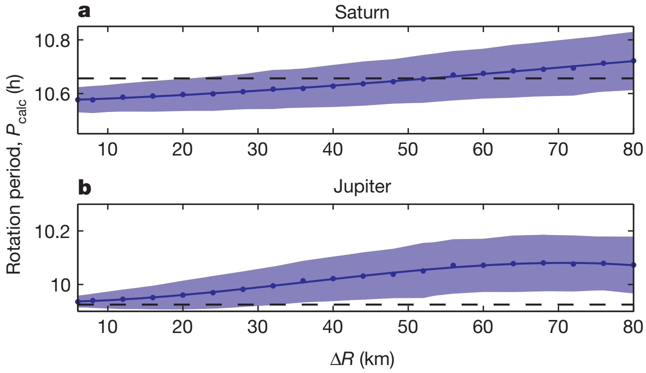


Figure 4 | Solutions for the rotation periods of Saturn and Jupiter when assuming improved gravity data. **a**, Saturn; **b**, Jupiter. Shown are P_{calc} and its 1σ standard deviation (blue shading) when setting $\Delta J_{2n} = 10^{-9}$ and $\Delta P = 0.5$ h. The calculated period is given versus the assumed uncertainty in the observed mean radius R_{obs}.

P_{calc} decreases by ~15%. We therefore conclude that the future measurements by Cassini could be important to further constrain Saturn's rotation period.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 September 2014; accepted 2 February 2015.
Published online 25 March 2015.

1. Sterenborg, M. G. & Bloxham, J. Can Cassini magnetic field measurements be used to find the rotation period of Saturn's interior? *Geophys. Res. Lett.* **37,** 11201 (2010).
2. Smith, B. A. *et al.* A new look at the Saturn system: the Voyager 2 images. *Science* **215,** 504–537 (1982).
3. Gurnett, D. A. *et al.* Radio and plasma wave observations at Saturn from Cassini's approach and first orbit. *Science* **307,** 1255–1259 (2005).
4. Gurnett, D. A. *et al.* The variable rotation period of the inner region of Saturn's plasma disk. *Science* **316,** 442–445 (2007).
5. Giampieri, G., Dougherty, M. K., Smith, E. J. & Russell, C. T. A regular period for Saturn's magnetic field that may track its internal rotation. *Nature* **441,** 62–64 (2006).
6. Sanchez-Lavega, A., Rojas, J. F. & Sada, P. V. Saturn's zonal winds at cloud level. *Icarus* **147,** 405–420 (2000).
7. Lindal, G. F., Sweetnam, D. N. & Eshleman, V. R. The atmosphere of Saturn—an analysis of the Voyager radio occultation measurements. *Astrophys. J.* **90,** 1136–1146 (1985).
8. Anderson, J. D. & Schubert, G. Saturn's gravitational field, internal rotation, and interior structure. *Science* **317,** 1384–1387 (2007).
9. Read, P. L., Dowling, T. E. & Schubert, G. Saturn's rotation period from its atmospheric planetary-wave configuration. *Nature* **460,** 608–610 (2009).
10. Zharkov, V. N. & Trubitsyn, V. P. *Physics of Planetary Interiors* 388 (Pachart Publishing House, 1978).
11. Hubbard, W. B. *Planetary Interiors* 1–343 (Van Nostrand Reinhold, 1984).
12. Schubert, G., Anderson, J., Zhang, K., Kong, D. & Helled, R. Shapes and gravitational fields of rotating two-layer Maclaurin ellipsoids: application to planets and satellites. *Phys. Earth Planet. Inter.* **187,** 364–379 (2011).
13. Kaspi, Y., Showman, A. P., Hubbard, W. B., Aharonson, O. & Helled, R. Atmospheric confinement of jet streams on Uranus and Neptune. *Nature* **497,** 344–347 (2013).
14. Jacobson, R. A. *JUP230 Orbit Solutions* http://ssd.jpl.nasa.gov/ (2003).
15. Jacobson, R. A. *et al.* The gravity field of the Saturnian system from satellite observations and spacecraft tracking data. *Astrophys. J.* **132,** 2520–2526 (2006).
16. Hubbard, W. B. *et al.* Structure of Saturn's mesosphere from the 28 SGR occultations. *Icarus* **130,** 404–425 (1997).
17. Flasar, F., Schinder, P. J., French, R. G., Marouf, E. A. & Kliore, A. J. Saturn's shape from Cassini radio occultations. *AGU Fall Meet. Abstr.* B8 (2013).
18. Helled, R., Schubert, G. & Anderson, J. D. Jupiter and Saturn rotation periods. *Planet. Space Sci.* **57,** 1467–1473 (2009).
19. Helled, R. Jupiter's occultation radii: implications for its internal dynamics. *Geophys. Res. Lett.* **38,** 8204 (2011).
20. Helled, R. & Guillot, T. Interior models of Saturn: including the uncertainties in shape and rotation. *Astrophys. J.* **767,** 113 (2013).
21. Guillot, T. The interiors of giant planets: Models and outstanding questions. *Annu. Rev. Earth Planet. Sci.* **33,** 493–530 (2005).
22. Fortney, J. J. & Nettelmann, N. The interior structure, composition, and evolution of giant planets. *Space Sci. Rev.* **152,** 423–447 (2010).
23. Dowling, T. E. Saturn's longitude: rise of the second branch of shear-stability theory and fall of the first. *Int. J. Mod. Phys. D* **23,** 1430006–86 (2014).
24. Higgins, C. A., Carr, T. D. & Reyes, F. A new determination of Jupiter's radio rotation period. *Geophys. Res. Lett.* **23,** 2653–2656 (1996).
25. Porco, C. C. *et al.* Cassini imaging of Jupiter's atmosphere, satellites and rings. *Science* **299,** 1541–1547 (2003).
26. Iess, L., Finocchiaro, S. & Racioppa, P. The determination of Jupiter and Saturn gravity fields from radio tracking of the Juno and Cassini spacecraft. *AGU Fall Meet. Abstr.* B1 (2013).
27. Finocchiaro, S. & Iess, L. in *Spaceflight Mechanics 2010* Vol. 136, 1417–1426 (American Astronautical Society, 2010).

Acknowledgements We thank G. Schubert, M. Podolak, J. Anderson, L. Bary-Soroker, and the Juno science team for discussions and suggestions. We acknowledge support from the Israel Space Agency under grants 3-11485 (R.H.) and 3-11481 (Y.K.).

Author Contributions R.H. led the research. R.H. and Y.K. initiated the research and wrote the paper. E.G. designed the optimization approach and executed all the calculations. R.H. computed interior models and defined the parameter space for the figure functions. All authors contributed to the analysis and interpretation of the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.H. (rhelled@post.tau.ac.il).

METHODS

The theory of figures. The theory of figures was first introduced by Clairaut[28], who derived an integro-differential equation for calculating the oblateness of a rotating planet in hydrostatic equilibrium with a non-uniform density profile. The method was further developed by Zharkov and Trubitsyn[10], who presented a theoretical description to connect the density profile of a hydrostatic planet with its gravitational moments J_{2n}, extending the theory to an arbitrary order. The basic idea of the method is that the density profile of a rotating planet in hydrostatic equilibrium can be derived by defining the layers as level surfaces, that is, surfaces of a constant potential (called the effective potential) that is set to be the sum of the gravitational potential and the centrifugal potential[10,11]:

$$U=\frac{GM}{r}\left(1-\sum_{n=1}^{\infty}\left(\frac{a}{r}\right)^{2n}J_{2n}P_{2n}(\cos\theta)\right)+\frac{1}{2}\omega^2r^2\sin^2\theta \quad (1)$$

where r is the radial distance, a is the equatorial radius of the geoid, GM is its mass multiplied by the gravitational constant, θ is the colatitude, and ω is the angular velocity given by $2\pi/P$, with P being the rotation period.

The internal density profile and the gravitational moments are linked through the smallness parameter $m=\omega^2R^3/GM$, where R is the mean radius of the planet. The gravitational moments can be expanded as a function of m by:

$$J_2=ma_{2,1}+m^2a_{2,2}+m^3a_{2,3} \quad (2a)$$

$$J_4=m^2a_{4,2}+m^3a_{4,3} \quad (2b)$$

$$J_6=m^3a_{6,3} \quad (2c)$$

where $a_{2n,k}$ are the expansion coefficients in smallness parameter. As $J_2 \gg J_4 \gg J_6$ higher-order coefficients correspond to a higher-order expansion in m. The gravitational moments J_{2n} are determined from the combination of the internal density distribution as well as the rotation period. As a result, unless the density profile of Saturn (or any other giant planet) is perfectly known there is no simple way to derive the rotation period and vice versa.

For the investigation of planetary figures, the equation for level surfaces can be written in the form of a spheroid that is a generalized rotating ellipsoid. Then, the planetary radius r at every latitude can be expressed as a function of the polar angle θ (colatitude), and the flattening parameters f, k and h by[10,12]:

$$r(\theta)=a\left[1-f\cos^2\theta-\left(\frac{3}{8}f^2+k\right)\sin^2 2\theta+\frac{1}{4}\left(\frac{1}{2}f^3+h\right)\left(1-5\sin^2\theta\right)\sin^2 2\theta\right] \quad (3)$$

where $f=(a-b)/a$ is the flattening (with b being the polar radius), and k and h are the second-order and third-order corrections, respectively[10]. While f is strictly the flattening of the object, k and h represent the departure of the level-surface from a precise rotating ellipsoid to second and third order in smallness parameter, and their values are expected to be much smaller than f.

To third order, the three flattening parameters f, k, h at the planetary surface (the effective potential surface) can be written as a sum of figure functions defined by[12]:

$$f=mF_1+m^2F_2+m^3F_3 \quad (4a)$$

$$k=m^2K_2+m^3K_3 \quad (4b)$$

$$h=m^3H_3 \quad (4c)$$

Finally, using the relation between the first three even gravitational moments and the figure functions for a density profile that is represented by a 6th-order polynomial[8,12,13] and by applying the theory of figures as a set of differential equations, the gravitational moments and the figure functions can be related as power series in the small rotational parameter m (see equation (72) in ref. 12). Since only J_2, J_4 and J_6 are currently known for Saturn (and Jupiter) we expand only up to third order in m. Although higher-order harmonics are not expected to be zero, the corrections will be $O(m^4)$ and therefore their contribution will be small.

The optimization method. Since the flattening parameters (and figure functions) depend on the density distribution, which is unknown, we take a general approach that is designed to relate the planetary rotation period to its gravitational field without putting tight constraints on the internal structure. We therefore develop an optimization method that searches for the solutions that reproduce Saturn's measured gravitational field within the widest possible pre-defined parameter space. The figure functions (F, H, K) are allowed to vary over their widest possible physical range, and the smallness parameter m is allowed to vary within a range that reflects the uncertainty in the rotation period P. A solution for these parameters is sought while meeting the requirement that Saturn's measured physical properties are reproduced.

First, an optimization function is defined as the sum of the normalized absolute differences between the observed moments and the calculated moments and is given by:

$$Y=\sum\left(\frac{\left|J_2-J_2^{\mathrm{obs}}\right|}{\left|\Delta J_2^{\mathrm{obs}}\right|}+\frac{\left|J_4-J_4^{\mathrm{obs}}\right|}{\left|\Delta J_4^{\mathrm{obs}}\right|}+\frac{\left|J_6-J_6^{\mathrm{obs}}\right|}{\left|\Delta J_6^{\mathrm{obs}}\right|}\right) \quad (5)$$

where J_2, J_4 and J_6 are the gravitational moments calculated using equations (2a–c), J_2^{obs}, J_4^{obs} and J_6^{obs} are the measured gravitational moments, and $\Delta J_2^{\mathrm{obs}}$, $\Delta J_4^{\mathrm{obs}}$ and $\Delta J_6^{\mathrm{obs}}$ are the uncertainties on the measured gravitational moments[14,15]. Since the observations include only the first three even harmonics everything is calculated to third order, but the method can be modified to include higher-order terms. The data that are used by the model are summarized in Extended Data Table 1.

Next, we minimize the optimization function Y with respect to the control variables F_1, F_2, F_3, K_2, K_3, H_3 and m, that is, the figure functions, and the smallness parameter. Starting from an arbitrary initial guess for each of the seven control variables (within the predefined limits), a solution is sought such that the optimization function reaches a minimum. Several nonlinear constraints are imposed while searching for the solution. First, we require that the difference between each calculated and the measured gravitational moments must be smaller than the uncertainty of the measurement error, that is, $\left|J_2-J_2^{\mathrm{obs}}\right|-\left|\Delta J_2^{\mathrm{obs}}\right|<0$, $\left|J_4-J_4^{\mathrm{obs}}\right|-\left|\Delta J_4^{\mathrm{obs}}\right|<0$, and $\left|J_6-J_6^{\mathrm{obs}}\right|-\left|\Delta J_6^{\mathrm{obs}}\right|<0$. Note that this requirement is additional to the minimization of Y since we ask that not only the overall difference between the observed and calculated gravitational moments is minimized, but that individually, each of the calculated moments stays within the uncertainty of its observed counterpart.

The parameter f is the planetary flattening; as a result, f must be a small positive number (for Saturn $f\approx 0.1$). The second- and third-order corrections, k and h, are substantially smaller than f, but could be either positive or negative. Thus, in order to keep our calculation as general as possible we allow the three flattening parameters to vary between their maximum physical values, -1 and 1. In Extended Data Fig. 1 we show the calculated values for f, h, k for Saturn for the case in which the figure functions are not constrained and ΔR is taken to be 50 km. f is found to be of the order of 0.1, consistent with the measured flattening of Saturn[7,20], while the second-order and third-order corrections are found to be of the order of 10^{-3}. As F_1 is the first-order expansion for f, we have $f-F_1m=O(m^2)$, meaning that for Saturn $|F_1|<1$. Similarly expanding recursively the other coefficients of f, and also those of k and h, implies that all figure functions are bound between -1 and 1. To keep our calculation as general as possible we allow all the figure functions to vary between -1 and 1. The solution though, which must also fit the gravitational field, constrains the flattening parameters and figure functions to a much narrower range. The solution is derived by using a numerical algorithm that is designed to solve constrained nonlinear multivariable functions. We use a sequential quadratic method that formulates the above nonlinear constraints as Lagrange multipliers[29]. The optimization is completed once the tolerance values for the function (10^{-3}) and the constraints (10^{-12}) are met.

A single optimization would be sufficient if the problem was well defined. In such a case, there would have been a unique solution that is independent of the initial guess. However, since in our case there are only three equations (equations (2a–c)) and seven unknowns (six figure functions and the smallness parameter), the problem is inherently ill-defined and therefore has no unique solution. Nevertheless, we can still reach a solution using a 'statistical' approach in which we repeat the optimization process enough times to achieve a statistically stable solution. In each case, the initial guess of the various parameters is chosen randomly within the defined bounds of each parameter. A statistical significance is reached when we repeat the optimization 2,000 times (verified with 10^4 optimizations). We can then use the solutions from the 2,000 optimizations to compute the mean value and its standard deviation for each variable. An example for a specific case is given in Extended Data Fig. 2, where the solutions for the gravitational moments (Extended Data Fig. 2a–c) are distributed around the mean value, and the distribution of the solutions for the figure functions (Extended Data Fig. 2d–j) has a large range. From each such experiment we eventually calculate two numbers: the mean rotation period P_{calc} and its standard deviation.

Expected improvements from Cassini's proximal orbits. Another objective of this research is to determine whether the improved gravity measurements of the low-order gravitational moments (J_2, J_4, J_6) by the Cassini's proximal orbits mission can be used to better constrain Saturn's rotation period. The data are not yet available, but are expected to be within the current uncertainty, so all we can do at present is estimate the rotation period and its standard deviation when the uncertainty on the gravitational moments (ΔJ_{2n}) is of the order of 10^{-9} while using the currently known values of J_2, J_4, and J_6. The result for this exercise is presented in Fig. 4. It is found that for Saturn this yields a 15% improvement in the derived

standard deviation of its rotation period. However, it is important to remember that the true values of the gravitational moments can be any value within the current uncertainty. We find that the order of magnitude of the standard deviation is not very sensitive to the actual value of J_{2n} but is more affected by the allowed uncertainty (that is, ΔJ_{2n}); we can therefore conclude that the Cassini proximal orbit measurement is useful to further constrain Saturn's rotation period.

Accounting for the planetary shape. Our optimization method can include additional constraints. Since the planetary shape could be used to constrain the rotation period[18,20], we also run cases in which we account for Saturn's shape (see equation (4)). Then the optimization includes the constraint that the calculated mean radius R should be consistent with the mean radius that is inferred from measurements. Thus, the calculated mean radius R_{calc} should be less than a specified uncertainty, that is:

$$|R_{\mathrm{calc}} - R_{\mathrm{obs}}| - |\Delta R| < 0 \quad (6)$$

where R_{obs} is the mean radius estimated from measurements of the planetary shape, and ΔR is the uncertainty associated with the measured radius. In the standard case we set this uncertainty to be 40 km, which is large compared to the measured uncertainty in Saturn's shape[7]. This provides a fourth equation to our optimization method and allows a considerable reduction in the rotation period uncertainty (Figs 2b and 3b).

Although Saturn's measured shape (radius as a function of latitude) is well determined from occultation measurements[7,16], one should note that there is a difference between the measurement uncertainty (estimated to be ~6 km; refs 7, 17) and the actual uncertainty (of the order of a few tens of kilometres[20]). The actual uncertainty is relatively large because the planet's measured shape is also affected by atmospheric winds, which distort the hydrostatic shape. The equatorial region of Saturn is affected by the large equatorial winds, and indeed the dynamical heights of the equator are found to be ~120 km (refs 7, 8, 16 and 20). On the other hand, the polar region is less affected by winds, and therefore the polar radii better reflect Saturn's hydrostatic shape. There are, however, no available occultation measurements of Saturn's polar regions. In addition, Saturn's north–south asymmetry in wind structure introduces an additional uncertainty in determining its shape. As a result, a more conservative uncertainty in Saturn's mean radius is estimated to be ~40 km (ref. 20).

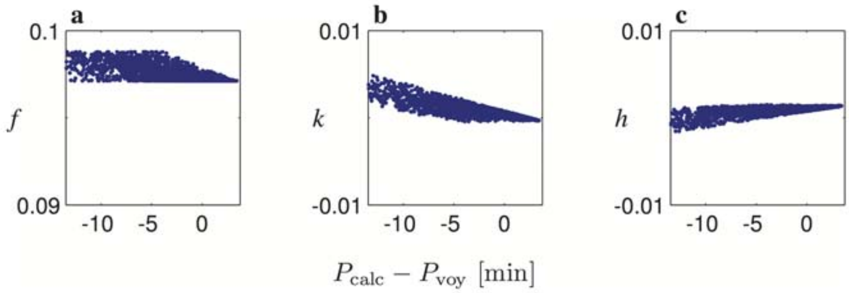
Interestingly, the solution for the rotation period for the case without the shape constraint does not necessarily contain the solution when the constraint on the shape is included. This is caused by the fact that taking into account only the gravitational moments, for Saturn, leads to a solution with relatively long rotation periods, while the measured shape pushes to shorter rotation periods. This effect is illustrated in Extended Data Fig. 3 where the solutions for the rotation period are shown in the phase-space of the constraints for ΔP and ΔR. When there is no constraint on the shape, and ΔP is large (upper-right, red region), the solution converges to a relatively long rotation period; as ΔP decreases, solutions with shorter rotation periods can be found (upper-left, blue region, see also Fig. 2a). When the constraint on the shape is included, even when the range of the rotation period is large (bottom-right, blue region), the solution converges into a short rotation period. The dashed line shows the transition between the regime where the constraint on the period (above the dashed line) to the regime where the constraint on the shape is more important. For the physical region we are interested in, as the constraint on the shape is increased (as $\Delta \bar{R}$ decreases), it becomes more dominant than the constraint on the rotation period, leading to a shorter rotation period outside the range of the solution without the shape constraint.

Constraining the figure functions. We also present results for cases in which the figure functions are constrained as well (Figs 2c and 3c for Saturn and Jupiter, respectively). In these cases, the figure functions are limited to a range that is determined from realistic interior models[21,22]. To put limits on their values, we run two limiting interior models for both Saturn and Jupiter and derive the values of the figure functions. The first case is one of a massive core, for which we assume a constant-density core with a density of $\sim 1.5 \times 10^4\,\mathrm{kg\,m^{-3}}$, reaching 20% of the planet's radius. In the second case, the density is continuous with no core and is represented by a 6th-order polynomial. For this case, the first-degree term of the polynomial is missing so that the derivative of the density goes to zero at the centre. Another constraint sets this value to zero at the core-envelope boundary for models with cores. We then use the derived values of the figure functions for each case to limit the values of the variables F_1, F_2, F_3, K_2, K_3 and H_3.

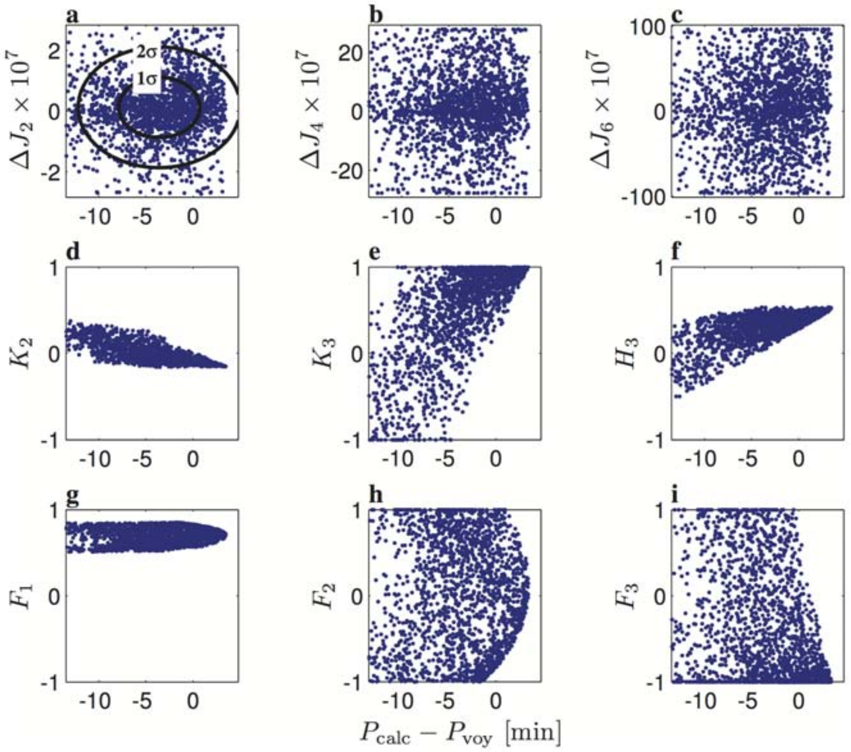
The figure functions we derive for the massive core and continuous density profile cases for Saturn and Jupiter are summarized in Extended Data Table 2. The density profiles for Saturn (top) and Jupiter (bottom) are shown in Extended Data Fig. 4. The values of the figure functions for interior models intermediate to these extreme cases all lie within the values of the extreme cases, implying that we have taken an exclusive set of values. To make sure that we account for a relatively large range of possible interior models, even within this fairly constrained case, we allow the figure function values to vary around the average value between the two models by a factor of two of the difference. The results (for Saturn) are shown in Extended Data Fig. 5. It is clear that in this case, due to the limitation on the figure functions, the parameter space of possible solutions is smaller, allowing a more accurate determination of P_{calc}. This range still accounts for a large variation in Saturn's density profile. While there is an improvement in the determination of P_{calc} when the shape and figure functions are tightly constrained, there is also a clear advantage in keeping the method as general as possible. The inferred result is then not associated with a specific interior model and/or does not rely on shape measurements. Our method is therefore also useful for estimating the rotation period of giant planets with less accurate determinations of their physical properties. For the icy planets, Uranus and Neptune, only J_2 and J_4 are currently known, yet a simplified version of this optimization (to second order) can be applied and gives rotation periods within 2% of the Voyager radio periods, allowing an independent method for estimating their rotation periods from their gravitational moments. Furthermore, our method could also be applied to derive the rotation periods of exoplanets for which the gravitational moments can be estimated[30,31].

Code availability. The optimization code (written in Matlab) that was used to calculate the rotation period and its standard deviation is available at http://www.weizmann.ac.il/EPS/People/Galanti/research.

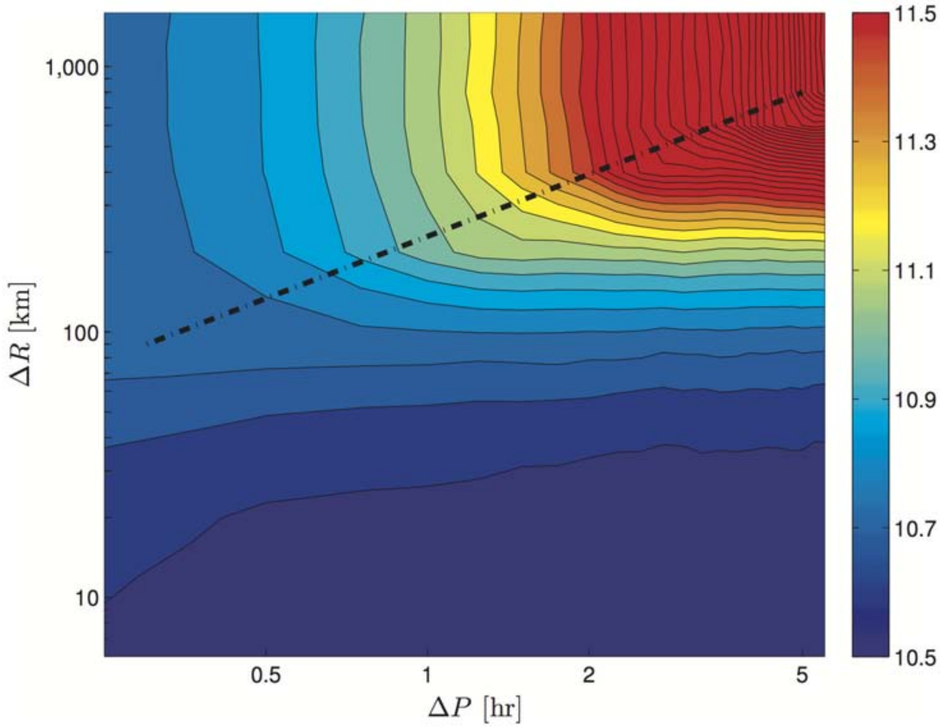
28. Clairaut, A. C. *Traite de la Figure de la Terre, tiree des Principes de l'Hydrostatique* (Paris Courcier, 1743).
29. Nocedal, J. & Wright, S. J. *Conjugate Gradient Methods* 102–120 (Springer, 2006).
30. Carter, J. A. & Winn, J. N. Empirical constraints on the oblateness of an exoplanet. *Astrophys. J.* **709,** 1219–1229 (2010).
31. Kramm, U., Nettelmann, N., Fortney, J. J., Neuhäuser, R. & Redmer, R. Constraining the interior of extrasolar giant planets with the tidal Love number k_2 using the example of HAT-P-13b. *Astron. Astrophys.* **538,** A146 (2012).



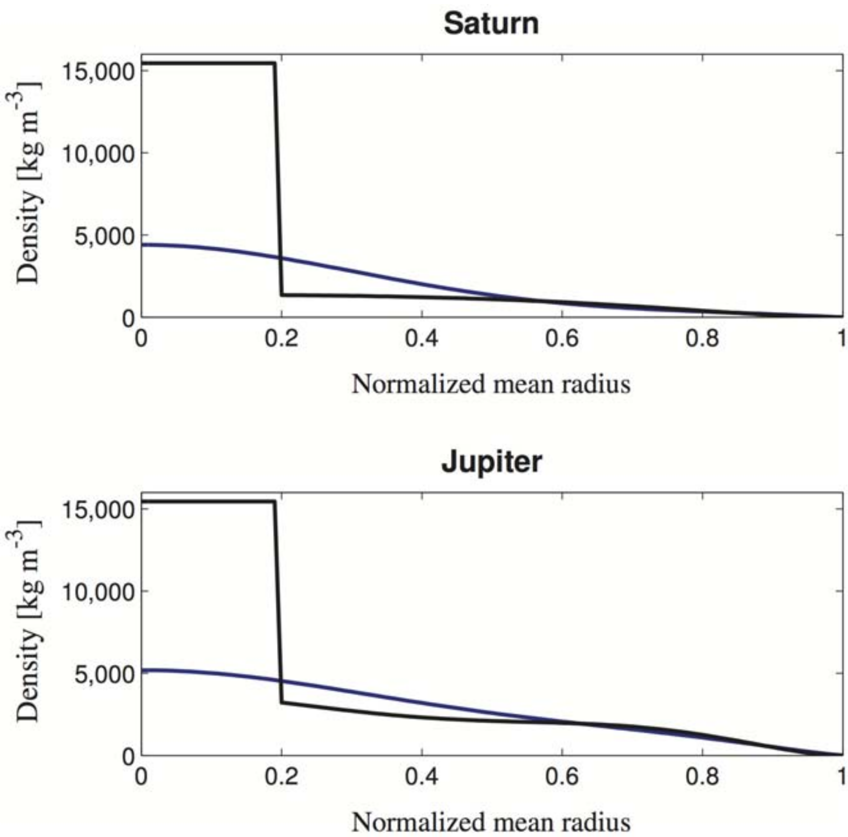
Extended Data Figure 1 | The flattening parameters calculated by the model without constraining the figure functions. Shown are f, k, h for Saturn with $\Delta R = 50\,\mathrm{km}$.



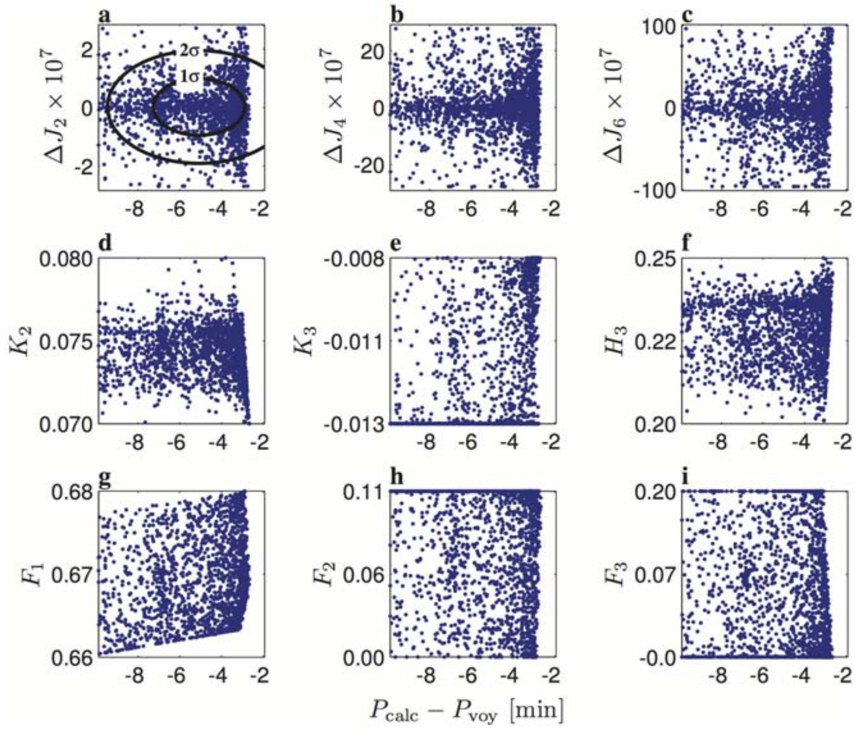
Extended Data Figure 2 | An example of our statistical optimization model for deriving the rotation period. The results are shown for a case for which the range of allowed rotation period is between 10 h 24 min and 10 h 54 min. The solution is based on a combination of 2,000 individual sub-cases, each of them representing a case with specific random initial conditions within the defined parameter space. **a**–**c**, A scatter plot (similar to Fig. 1a) of the distribution of solutions on the plane of the calculated rotation period P_{calc} minus P_{voy} versus each of ΔJ_2, ΔJ_4 and ΔJ_6. Each blue dot represents one sub-case converged solution. In **a** the inner and outer black circles show the first and second standard deviations, respectively. **d**–**j**, The distribution of solutions for the figure functions K_2, K_3, H_3, F_1, F_2 and F_3, respectively.



Extended Data Figure 3 | Saturn's calculated rotation period versus the uncertainty in the assumed rotation period and radius. P_{calc} shown (colour scale) as a function of ΔP (h) and ΔR (km). The dashed line presents the transition from the regime where the constraint on the rotation period (above the dashed line) to the regime where the constraint on the shape is more dominant.



Extended Data Figure 4 | Radial density profiles for two different interior models for Saturn (top) and Jupiter (bottom). The black curves correspond to models with very large cores and the blue curves are no-core models in which the density profile is represented by 6th-order polynomials. For the massive-core case we assume a constant core density of $\sim 1.5 \times 10^4 \mathrm{kg\,m^{-3}}$, reaching 20% of the planet's radius. The density profiles are constrained to match the planetary mass, J_2, J_4, J_6, mean radius, and the atmospheric density and its derivative at 1 bar (see details in refs 8, 13 and 18). We then use the difference in the values of the figure functions in the two limiting cases to limit their values.



Extended Data Figure 5 | The calculated flattening parameters when the figure functions are limited by interior models. **a**–**c**, A scatter plot (similar to Fig. 1a) of the distribution of solutions on the plane of the calculated rotation period $P_{\rm calc}$ minus $P_{\rm voy}$ versus ΔJ_2, ΔJ_4 and ΔJ_6, respectively. Each blue dot represents one sub-case converged solution. The calculated mean radius was set to be within 20 km of Saturn's observed mean radius. In **a** the inner and outer black circles show the first and second standard deviations, respectively. **d**–**j**, The distribution of solutions for the figure functions K_2, K_3, H_3, F_1, F_2 and F_3, respectively.

Extended Data Table 1 | The physical properties of Saturn and Jupiter used in the analysis

	Saturn	Jupiter
Mass (10^{24} kg)	568.36	1,898.3
Rotation Period (System III)	10hr 39m 24s	9hr 55m 29s
Mean Radius (km)	58,232	69,911
J_2 (10^{-6})	$16,290.71 \pm 0.27$	$14,696.43 \pm 0.21$
J_4 (10^{-6})	-935.83 ± 2.77	-587.14 ± 1.68
J_6 (10^{-6})	86.14 ± 9.64	34.25 ± 5.22

Data are taken from http://ssd.jpl.nasa.gov/?gravity_fields_op (JUP 230). The gravitational moments correspond to a reference equatorial radius of 60,330 km and 71,492 km for Saturn and Jupiter, respectively.

Extended Data Table 2 | The calculated figure functions based on interior models of Saturn and Jupiter

	Saturn - Massive Core	Saturn - No Core
F_1	0.66357	0.67720
F_2	0.07501	0.04743
F_3	0.043517	0.10790
K_2	0.22691	0.07535
K_3	-0.01039	-0.01174
H_3	0.10074	0.23489
	Jupiter - Massive Core	**Jupiter - No Core**
F_1	0.77014	0.76965
F_2	0.08512	0.08499
F_3	0.14874	0.15137
K_2	0.05868	0.05861
K_3	-0.00705	-0.01068
H_3	0.19885	0.20570

The values of the figure functions are derived for the two limiting cases of massive core and no-core (continuous density profile) for Saturn and Jupiter.

doi:10.1038/nature14341

Four-wave mixing experiments with extreme ultraviolet transient gratings

F. Bencivenga[1], R. Cucini[1], F. Capotondi[1], A. Battistoni[1,2], R. Mincigrucci[1,3], E. Giangrisostomi[1,2], A. Gessini[1], M. Manfredda[1], I. P. Nikolov[1], E. Pedersoli[1], E. Principi[1], C. Svetina[1,2], P. Parisse[1], F. Casolari[1], M. B. Danailov[1], M. Kiskinova[1] & C. Masciovecchio[1]

Four-wave mixing (FWM) processes, based on third-order nonlinear light–matter interactions, can combine ultrafast time resolution with energy and wavevector selectivity, and enable the exploration of dynamics inaccessible by linear methods[1–7]. The coherent and multi-wave nature of the FWM approach has been crucial in the development of advanced technologies, such as silicon photonics[8], subwavelength imaging[9] and quantum communications[10]. All these technologies operate at optical wavelengths, which limits the spatial resolution and does not allow the probing of excitations with energy in the electronvolt range. Extension to shorter wavelengths—that is, the extreme ultraviolet and soft-X-ray ranges—would allow the spatial resolution to be improved and the excitation energy range to be expanded, as well as enabling elemental selectivity to be achieved by exploiting core resonances[5–7,11–14]. So far, FWM applications at such wavelengths have been prevented by the absence of coherent sources of sufficient brightness and of suitable experimental set-ups. Here we show how transient gratings, generated by the interference of coherent extreme-ultraviolet pulses delivered by the FERMI free-electron laser[15], can be used to stimulate FWM processes at suboptical wavelengths. Furthermore, we have demonstrated the possibility of observing the time evolution of the FWM signal, which shows the dynamics of coherent excitations as molecular vibrations. This result opens the way to FWM with nanometre spatial resolution and elemental selectivity, which, for example, would enable the investigation of charge-transfer dynamics[5–7]. The theoretical possibility of realizing these applications has already stimulated ongoing developments of free-electron lasers[16–20]: our results show that FWM at suboptical wavelengths is feasible, and we hope that they will enable advances in present and future photon sources.

In FWM, three coherent fields interact within the sample via the third-order susceptibility $\chi^{(3)}$, which, unlike the second-order susceptibility, does not vanish by reason of the sample symmetry[1,2]; FWM processes thus represent the lowest-order (non-vanishing) optical nonlinearity for most materials. The FWM concept is the basis of many coherent nonlinear methods, such as coherent Raman scattering, multi-dimensional spectroscopy, and impulsive stimulated Rayleigh, Brillouin and Raman scattering[1–7,21]. These applications are based on dynamic (transient) gratings, generated by periodic modulations of the sample's optical properties by different excitations that ultimately drive the time evolution of the created gratings. Hence the capability to generate, control and probe the time dependence of extreme-ultraviolet (EUV) transient gratings at ultrafast timescales is an essential step towards the development of FWM at suboptical wavelengths. Though the potential high impact of using short-wavelength photons has been discussed in a number of theoretical and perspective works[7–10,12,21] and evidence for basic EUV/X-ray nonlinear processes has been reported[22–25], the experimental investigation of FWM in this spectral range has been prevented by the lack of fully coherent sources with high brilliance. In this Letter we report experimental evidence of FWM processes stimulated by EUV transient gratings, generated by two coherent EUV pulses provided by the FERMI FEL[15].

Figure 1a shows a sketch of the experiment: two FEL pulses (wavelength $\lambda_{EUV} = 27.6$ nm, estimated time duration 60–80 fs, pulse energy $I_{EUV1} \approx I_{EUV2} \approx 5$ μJ, spot size ~0.04 mm^2) are crossed (angle: $2\theta = 6.16°$) on a vitreous SiO_2 sample with the surface oriented orthogonally to the bisector of the FEL beams. The interference of the two pulses generates EUV transient gratings with a spatial periodicity $L = \lambda_{EUV}/2\sin(\theta) \approx 256.8$ nm. An optical pulse ($\lambda_{opt} = 392.8$ nm, time duration ~100 fs, pulse energy $I_{opt} \approx 2$ μJ, spot size ~0.002 mm^2), coplanar with the FEL beams, was sent into the sample at an angle of incidence $\theta_B = 49.9°$. The polarization of the fields was orthogonal to the scattering plane. The chosen values of λ_{EUV}, λ_{opt}, 2θ and θ_B fulfil the phase matching condition (see Fig. 1b), which in the present transient grating experiment reduces to: $\lambda_{EUV}\sin(\theta_B) = \lambda_{opt}\sin(\theta)$. This determines the propagation direction of the FWM signal: $\boldsymbol{k}_{FWM} = \boldsymbol{k}_{EUV1} - \boldsymbol{k}_{EUV2} + \boldsymbol{k}_{opt}$, where $\boldsymbol{k}_{EUV1,2}$ and $\boldsymbol{k}_{opt}$ are the wavevectors of the EUV fields and of the optical field, respectively. Along $\boldsymbol{k}_{FWM}$ the nonlinear signal, radiated by different portions of the sample, adds coherently, yielding a FWM signal propagating downstream from the sample as a well-defined beam[1–7,21]. The experiment was carried out at the DiProI end-station using a purposely designed experimental set-up based on reflective optics (see Methods for details). Such a set-up lets us adjust 2θ, θ_B, the time delay between the crossed FEL pulses ($\Delta t_{EUV\text{-}EUV}$) and the delay between the FEL pulses and the optical pulse (Δt). FEL-pump/optical-probe transient reflectivity measurements were used to determine $\Delta t_{EUV\text{-}EUV} = 0$ and to equalize the fluence of the two FEL pulses in the

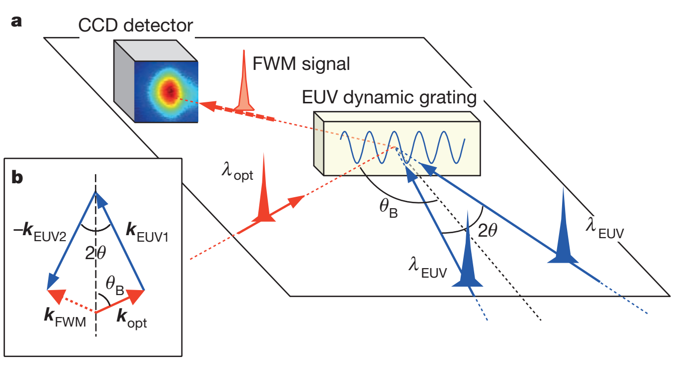


Figure 1 | FWM experiments with EUV transient gratings. **a**, Sketch of the FEL-based FWM experiment: $2\theta = 6.16°$, $\theta_B = 49.9°$, $\lambda_{EUV} = 27.6$ nm and $\lambda_{opt} = 392.8$ nm are the crossing angle between the two FEL beams that generate the EUV dynamic grating, the angle between their bisector (dotted black line) and the optical beam, the FEL wavelength and the laser wavelength, respectively. A CCD sensor is placed in the expected propagation direction of the FWM signal beam ($\boldsymbol{k}_{FWM}$), which is determined by the 'phase matching' (shown in **b**; here $\boldsymbol{k}_{opt}$, $\boldsymbol{k}_{EUV1}$ and $\boldsymbol{k}_{EUV2}$ are the wavevectors of the optical and of the two FEL pulses, respectively).

[1]Elettra-Sincrotrone Trieste, Strada Statale 14 km 163,5 in Area Science Park, I-34012 Basovizza, Trieste, Italy. [2]Dipartimento di Fisica, Università degli Studi di Trieste, Piazzale Europa, I-34127 Trieste, Italy. [3]Dipartimento di Fisica e Geologia, Università degli Studi di Perugia, Via Pascoli, I-06123 Perugia, Italy.

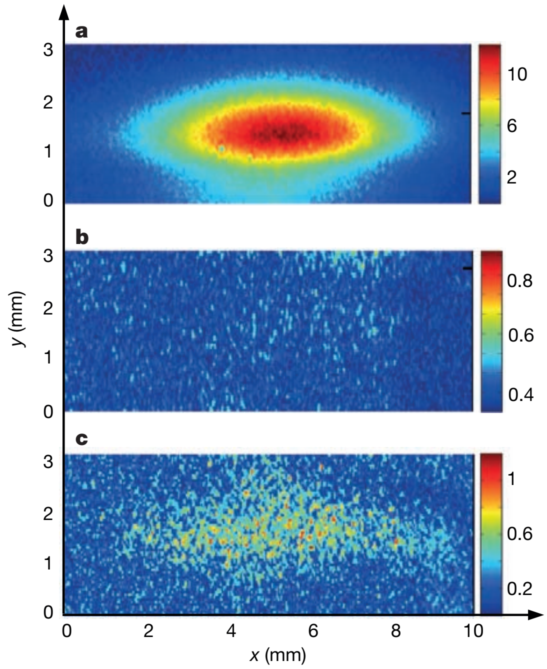


Figure 2 | FWM signal stimulated by EUV transient gratings. **a**–**c**, Images from the CCD placed in the direction of $\boldsymbol{k}_{FMW}$ (here x and y are the directions parallel and orthogonal to the scattering plane, respectively, while the colour scale reports the average number of photons per FEL shot). The images were acquired at $\Delta t = 0$ (**a**), $\Delta t = -0.5$ ps (**b**) and $\Delta t = 70$ ps (**c**).

interaction region. The FWM signal was detected by a CCD (charge-coupled device) placed along $\boldsymbol{k}_{FWM}$. After illumination by a few hundred FEL shots, creation of permanent gratings was observed at FEL fluences larger than $\sim$50 mJ cm^{-2} per pulse (see Methods for details). The illumination of these permanent gratings by the optical laser led to a diffracted signal along $\boldsymbol{k}_{FWM}$ even in the absence of the FEL pulses. During the FWM measurements the FEL fluence was kept below 50 mJ cm^{-2} in order to avoid the creation of permanent gratings.

Figure 2a shows a CCD image of the FWM signal beam acquired at $\Delta t = 0$, whereas Fig. 2b and c reports the same measurement for $\Delta t = -0.5$ ps and 70 ps, respectively. The observation of a signal that propagates along $\boldsymbol{k}_{FWM}$ for $\Delta t > 0$ in the concurrent presence of the three input beams undoubtedly demonstrates the occurrence of FWM processes stimulated by EUV transient gratings. The large angular spread observed along the plane of the input beams can be attributed to relaxed phase-matching conditions, probably due to the fact that both L and the EUV absorption length ($L_{EUV} \approx 30$ nm) are shorter than λ_{opt}. The total intensity of the FWM signal (I_{FWM}) at $\Delta t = 0$ was $\sim 6 \times 10^5$ photons per shot, corresponding to an efficiency $\eta_{FWM} = I_{FWM}/I_{opt} \approx 1.5 \times 10^{-7}$.

Concerning the magnitude of the nonlinear effect, the strength of the FWM field radiated by the sample (E_{FWM}) is related to those of the input fields as $E_{FWM} \propto \chi^{(3)} E_{EUV1} E_{EUV2} E_{opt}$ (refs 1–3, 5–7). In the present case we estimated $E_{EUV1} \approx E_{EUV2} \approx 0.8 \times 10^9$ V m^{-1}, so that an effective value for the third-order susceptibility $|\chi^{(3)*}| \approx (\eta_{FWM})^{1/2}/(E_{EUV1}E_{EUV2}) \approx 6 \times 10^{-22}$ m^2 V^{-2} (at $\Delta t = 0$) can be estimated. This value is within the expected order of magnitude, since $|\chi^{(n)}|$ scales as E_a^{1-n}, where $E_a \approx e/(4\pi\varepsilon_0 a_0^2) \approx 5 \times 10^{11}$ V m^{-1} is the atomic field strength (here e, ε_0 and a_0 are the elementary charge, the dielectric constant and the Bohr radius, respectively)[2,5,6,26].

The time evolution of η_{FWM} is shown in Fig. 3a. The data are scaled by the factor $\alpha = I_{EUV1}(\Delta t = 0)I_{EUV2}(\Delta t = 0)/[I_{EUV1}(\Delta t)I_{EUV2}(\Delta t)]$. Approaching $\Delta t = 0$ from negative delays, a fast rise of the FWM signal, compatible with a cross-correlation profile ($R_{cc}(\Delta t)$), was detected. For $\Delta t > 0$, an appreciable FWM signal is observed over the whole probed range of Δt (that is, up to 130 ps). The signal observed at large values of Δt (>10 ps) can be ascribed to thermal relaxation and longitudinal acoustic modes[3]. The latter are expected to induce a signal modulation at frequency $\nu_{LA} = c_s/L \approx 23.2$ GHz (where $c_s = 5{,}970$ m s^{-1} is the velocity of sound in vitreous SiO_2), which is compatible with our data. Figure 3b displays the FWM signal in the range 0 to 1.6 ps after $R_{cc}(\Delta t)$ has been subtracted. Here the time structure of the FWM signal is compatible with two oscillations at frequencies $\nu_1 \approx 1.15$ THz and $\nu_2 \approx 4.1$ THz. The latter frequency roughly matches that of the ν_{2b} Raman modes, due to tetrahedral bending, while the former, which accounts for the leading modulation of the observed FWM signal, can be ascribed to F_1 hyper-Raman modes, which involve rotations of SiO_4 tetrahedra (see Methods for details).

Our results demonstrate the generation of dynamic EUV gratings using fully coherent FEL pulses. The nonlinear interactions between the induced grating and an optical pulse allowed us to observe the first FEL-stimulated FWM signal, which encodes the dynamics of impulsively stimulated collective vibrational modes with wavevector $k_{ex} = |\boldsymbol{k}_{EUV1} - \boldsymbol{k}_{EUV2}|$ (ref. 3). This shows how EUV transient gratings can be used to drive coherent excitations into the sample, and paves the way for the practical exploitation of nonlinear optics in the EUV/soft X-ray (SXR) range. In this context, we aim to replace the optical pulse with an EUV/SXR pulse (EUV probing of optically stimulated transient gratings has already been reported[27,28]) to probe vibrational modes in the range $k_{ex} = 0.1$–1 nm^{-1} (ref. 21). This range, inaccessible by optical FWM (see Fig. 4a), can be probed by using photon wavelengths down to 10 nm (that is, well within the range of FERMI)[21], and is of special relevance for the investigation of disordered systems. Indeed, the origin of the thermodynamic peculiarities of such systems (for example, the highly debated excess specific heat) seems to be related to 'mesoscopic' heterogeneities that extend over a length scale of a few nanometres, which can be revealed by the anomalous behaviour of vibrational modes in the range $k_{ex} = 0.1$–1 nm^{-1} (refs 29, 30).

An opportunity offered by the FEL technology is the use of two-colour FEL pulses[16–19] to generate coherent populations of excited states

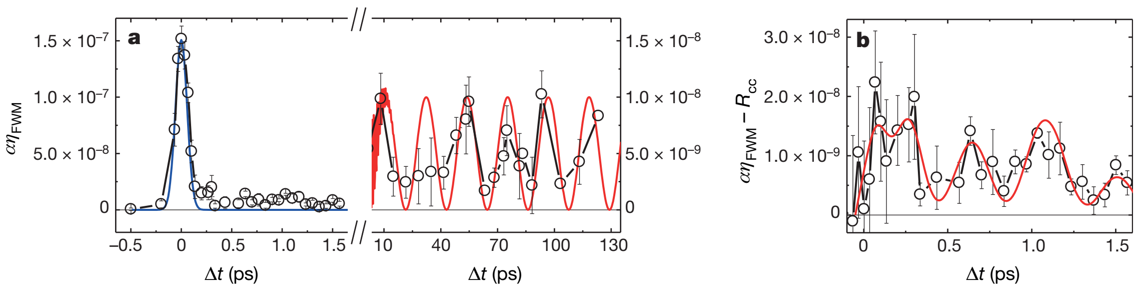


Figure 3 | Time evolution of the FWM signal. **a**, Black circles (connected by lines where there is enough distance between adjacent points) show the time dependence of the FWM signal, scaled to the intensity of the input beams (error bars are estimated as one standard deviation of the set of CCD images corresponding to the same Δt value); the blue and red lines are R_{cc} and the expected signal modulation due to acoustic modes, respectively (right-hand y axis plots same quantity as left-hand axis, but on an expanded scale). **b**, Black circles connected by lines are the FWM signal after R_{cc} is subtracted. The red line is the modulation due to oscillations at frequencies $\nu_1 = 1.15$ THz and $\nu_2 = 4.1$ THz. Error bars, as **a**.

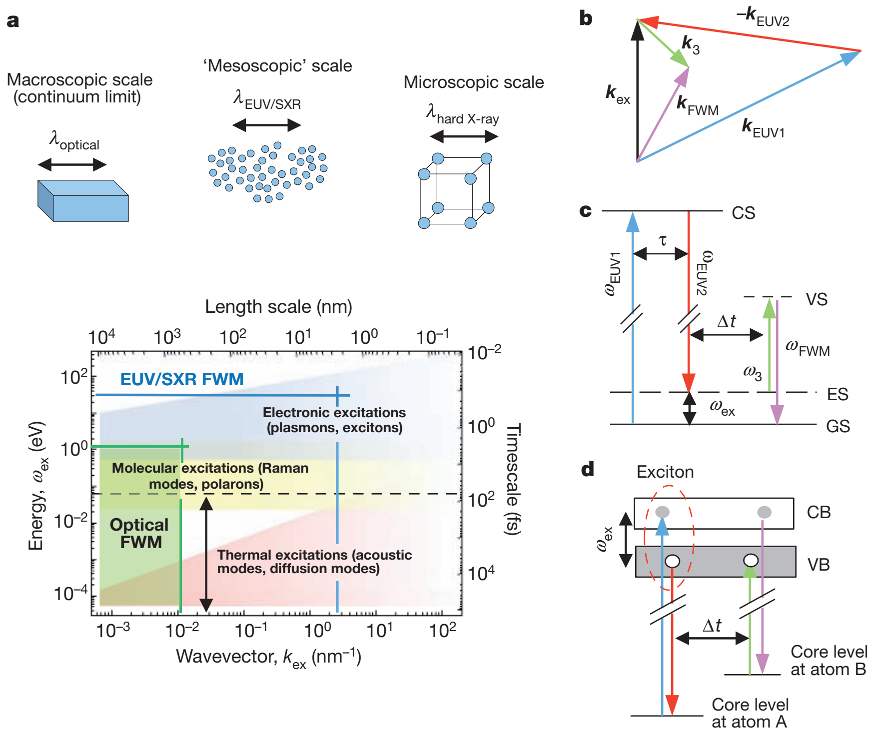


Figure 4 | Prospects for EUV/SXR FWM. a, Graph shows energy–wavevector (ω_{ex}–k_{ex}) range of typical excitations in condensed matter and corresponding timescales and length scales. The green area sketches the range accessible by optical FWM, while the double-ended black vertical arrow is the range probed in the present work; the horizontal dashed line is the time duration of the excitation pulses. The area delimited by the thick blue lines represents the (ω_{ex}–k_{ex}) range accessible by EUV/SXR FWM, with the addition of a EUV/SXR probe and the development of EUV/SXR CRS. Top row, sketches of materials probed by photons with wavelengths much longer than (left cartoon: continuum limit) and comparable to (right cartoon: microscopic scale) the interatomic distances; in between these two limits there is the almost unexplored mesoscopic range (middle cartoon), which can be probed by photons with wavelength comparable with some interatomic distances. **b**, Wavevector arrangement for a CRS process involving two-colour FEL excitation pulses ($|\mathbf{k}_{EUV1}| \neq |\mathbf{k}_{EUV2}|$); $\mathbf{k}_3$ and $\mathbf{k}_{FWM}$ are the wavevectors of the probe and signal beams, respectively. **c**, Level scheme in the case where ω_{EUV1} is tuned to the energy of a core state (CS) transition and $\omega_{EUV1} - \omega_{EUV2}$ to a lower-energy excitation (ω_{ex}). GS and ES are the ground and excited states, respectively, while VS is a virtual state (in standard CRS the CS is replaced by a VS); τ is the eventual time delay between the two excitation pulses; ω_3 and $\omega_{FWM} = \omega_3 + \omega_{ex}$ are the photon energies of the probe and signal beams, respectively. **d**, Independent control of the three EUV/SXR input beams might allow the energy of the excitation and probe beams to be separately tuned to CSs of two distinct atoms (atom A and atom B), hence allowing the excitation dynamics correlating two different atomic sites to be monitored[6,7] (a valence band exciton in this sketch[7]; here VB and CB are the valence and conduction bands, respectively).

via FWM processes, such as coherent Raman scattering (CRS)[2,4–7]. In CRS, the excitation energy and wavevector are set by the difference in photon energy ($\omega_{ex} = \omega_1 - \omega_2$) of the stimulating pulses and by the grating vector $\mathbf{k}_{ex} = \mathbf{k}_1 - \mathbf{k}_2$, respectively (see Fig. 4b and c). Compared to the optical regime, EUV/SXR pulses (10–1,000 eV) will allow us to coherently stimulate excitations at higher energy (for example, excitons with ω_{ex} = 1–10 eV)[5–7]. These can be probed by a third pulse, generated by the FEL[5,6,20] or by an external source[27,28]. Furthermore, when the field frequencies of the stimulating and/or probing pulses are tuned to core resonances of given atoms, the localization of core shells turns into an atomic-scale localization of the site at which the selected excitation is created and/or probed (see Fig. 4c and d). So by tuning the field frequencies and time delay it is possible to use FWM to gain real-time information on the excitation transfer between selected atomic sites. This possibility is forbidden in linear inelastic scattering, where the light–matter interaction takes place at the same atomic site[7].

In the future, multi-wave interactions would allow the application of a time delay (τ, see Fig. 4c) between the excitation pulses (for example, by setting $\Delta t_{EUV\text{-}EUV} = \tau \neq 0$ in our set-up). This is a substantial step towards multi-dimensional spectroscopy[4], and may enable resonant and non-resonant FWM signals to be distinguished. For experimental reasons, core-resonant FWM is still an unexplored field, though solid theoretical approaches and numerical studies in the SXR region are available[7,11]. Core-resonant FWM will soon be tested at FERMI, since many elements (such as C, Mg, Al, Si, Ti) have resonances within the photon energy range of this FEL (20–310 eV). The third harmonic of the FERMI FEL emission (photon energy range: 60–930 eV) and the forthcoming construction of self-seeded SXR FEL sources with a wider photon energy range (for example, SCLS-II, SwissFEL and PAL-XFEL) could most probably be used to exploit resonances at higher energy.

FWM with elemental selectivity could have applications[6,7] in the study of charge transfer dynamics in electrocatalytic processes occurring in photoelectrochemical cells; these devices mimic natural photosynthesis to convert sunlight into chemical energy. Photoelectrochemical cells are typically based on molecular complexes adsorbed on TiO_2 substrates, and have a conduction/valence band dominated by the Ti/O orbitals. FWM with elemental selectivity may be able to disentangle the timescales of electron and hole migration, by exploration of the Ti and O resonances, and to shed light on the water oxidation dynamics taking place at the catalyst by separating the signal originating from reaction centres[6]. Since elements of biological interest (C, N, O) have resonances in the EUV/SXR range, applications in life sciences are also very likely.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 November 2014; accepted 6 February 2015.

1. Bloembergen, N. Nonlinear optics and spectroscopy. *Rev. Mod. Phys.* **54,** 685–695 (1982).
2. Boyd, W. R. *Nonlinear Optics* (Academic, 2008).

3. Dhar, L., Rogers, J. A. & Nelson, K. A. Time-resolved vibrational spectroscopy in the impulsive limit. *Chem. Rev.* **94,** 157–193 (1994).
4. Cundiff, S. T. & Mukamel, S. Optical multidimensional coherent spectroscopy. *Phys. Today* **66,** 44–49 (2013).
5. Bencivenga, F. *et al.* Multi-colour pulses from seeded free-electron-lasers: towards the development of non-linear core-level coherent spectroscopies. *Faraday Discuss.* **171,** 487–503 (2014).
6. Bencivenga, F. *et al.* Nanoscale dynamics by short-wavelength four wave mixing experiments. *New J. Phys.* **15,** 123023 (2013).
7. Tanaka, S. & Mukamel, S. Coherent X-ray Raman spectroscopy: a nonlinear local probe for electronic excitations. *Phys. Rev. Lett.* **89,** 043001 (2002).
8. Foster, M. A. *et al.* Broad-band optical parametric gain on a silicon photonic chip. *Nature* **441,** 960–963 (2006).
9. Chen, P.-Y. & Al, A. Subwavelength imaging using phase-conjugating nonlinear nanoantenna arrays. *Nano Lett.* **11,** 5514–5518 (2011).
10. Boyer, V., Marino, A. M., Pooser, R. C. & Lett, P. D. Entangled images from four-wave mixing. *Science* **321,** 544–547 (2008).
11. Tanaka, S., Chernyak, V. & Mukamel, S. Time-resolved X-ray spectroscopies: nonlinear response functions and Liouville-space pathways. *Phys. Rev. A* **63,** 063405 (2001).
12. Adams, B. Nonlinear X-ray optics: the next phase for X-rays. *Nature Phys.* **7,** 675–676 (2011).
13. Glover, T. E. *et al.* X-ray and optical wave mixing. *Nature* **488,** 603–608 (2012).
14. Tamasaku, K., Sawada, K., Nishiboro, E. & Ishikawa, T. Visualizing the local optical response to extreme ultraviolet radiation with a resolution of $\lambda/380$. *Nature Phys.* **7,** 705–708 (2011).
15. Allaria, E. *et al.* Highly coherent and stable pulses from the FERMI seeded free-electron laser in the extreme ultraviolet. *Nature Photon.* **6,** 699–704 (2012).
16. Allaria, E. *et al.* Two-colour pump-probe experiments with a twin pulse-seed extreme ultraviolet free-electron laser. *Nature Commun.* **4,** 2476 (2013).
17. Lutman, A. A. *et al.* Experimental demonstration of femtosecond two-color X-ray free-electron-lasers. *Phys. Rev. Lett.* **110,** 134801 (2013).
18. Hara, T. *et al.* Two-colour hard X-ray free-electron laser with wide tunability. *Nature Commun.* **4,** 2919 (2013).
19. Lutman, A. A. *et al.* Demonstration of single-crystal self-seeded two-color X-ray free-electron lasers. *Phys. Rev. Lett.* **113,** 254801 (2014).
20. Marcus, G., Penn, G. & Zholents, A. A. Free-electron laser design for four-wave-mixing experiments with soft X-ray pulses. *Phys. Rev. Lett.* **113,** 024801 (2014).
21. Bencivenga, F. & Masciovecchio, C. FEL-based transient grating spectroscopy to investigate nanoscale dynamics. *Nucl. Instrum. Methods A* **606,** 785–789 (2009).
22. Shwartz, S. *et al.* X-ray second harmonic generation. *Phys. Rev. Lett.* **112,** 163901 (2014).
23. Rohringer, N. *et al.* Atomic inner-shell X-ray laser at 1.46 nanometres pumped by an X-ray free-electron laser. *Nature* **481,** 488–491 (2012).
24. Beye, M. *et al.* Stimulated X-ray emission for materials science. *Nature* **501,** 191–194 (2013).
25. Tamasaku, K. *et al.* X-ray two-photon absorption competing against single and sequential multiphoton processes. *Nature Phys.* **9,** 313–316 (2014).
26. Armstrong, J. A., Bloembergen, N., Ducuing, J. & Pershan, P. S. Interactions between light waves in a nonlinear dielectric. *Phys. Rev.* **127,** 1918–1939 (1962).
27. Tobey, R. I. *et al.* Transient grating measurement of surface acoustic waves in thin metal films with extreme ultraviolet radiation. *Appl. Phys. Lett.* **89,** 091108 (2006).
28. Sistrunk, E. F. *et al.* Extreme ultraviolet transient grating measurement of insulator-metal transition dynamics of VO_2. In *19th International Conference on Ultrafast Phenomena* http://dx.doi.org/10.1364/UP.2014.09.Wed.P3.44 (Paper 09.Wed.P3.44, Optical Society of America, 2014).
29. Schirmacher, W. *et al.* Vibrational excitations in systems with correlated disorder. *Phys. Status Solidi C* **5,** 862–866 (2008).
30. Ferrante, C. *et al.* Acoustic dynamics of network-forming glasses at mesoscopic wavelengths. *Nature Commun.* **4,** 1793 (2013).

Acknowledgements We acknowledge support from the Italian Ministry of University and Research through grants FIRB-RBAP045JF2 and FIRB-RBAP06AWK3, and from the European Research Council through grant 202804-TIMER. We thank M. Svandrlik and the FERMI-FEL team for support.

Author Contributions C.M. proposed and led the project to extend transient grating methods at suboptical wavelengths. F.B. conceived the experiment and coordinated all activities. F. Capotondi designed the set-up to split and recombine the FEL pulses. A.G. realized the set-up and, together with F. Capotondi, A.B. and R.M., integrated it into the end-station. F.B., R.C., A.B., A.G. and R.M. tested the set-up. I.P.N. and M.B.D. realized the set-up to control the optical pulse. F.B., R.C., F. Capotondi, A.B., R.M., E.G., M.M., E.P., E.P., C.S. and F. Casolari performed the experiment. A.B., R.M., F.B., R.C., F. Capotondi and M.M. carried out the data analysis. P.P., C.S., A.B. and R.M. performed the AFM measurements and analysis. F.B., R.C., F. Capotondi, A.B., M.B.D., M.K. and C.M. discussed the data. F.B. and C.M. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.B. (filippo.bencivenga@elettra.eu) or C.M. (claudio.masciovecchio@elettra.eu).

METHODS

Experimental set-up. The FEL focusing was provided by a Kirkpatrick-Baez active focusing system (located upstream of the DiProI end-station[31]), which was detuned to relax the spot size at the sample in order to avoid sample damage. The FEL photo-diagnostic data were acquired on a shot-to-shot basis by the Photon Analysis, Delivery and Reduction System (PADReS)[32], which connects the FEL source to the end-station; an Al filter placed along the photon transport line was used to remove the seed laser radiation.

The layout of the experimental set-up used to generate and probe EUV transient gratings is shown in Extended Data Fig. 1. The system to split and recombine the FEL beams (see Extended Data Fig. 1a) is based on three 70-mm-long carbon-coated mirrors (M1, M2 and M3) working at grazing incidence. It is mounted on a 580 mm $\times$ 280 mm baseplate and in the present case the grazing angles were $\sim 3^\circ$. M1 can be inserted into the path of the incoming FEL beam by a y-translation (see Extended Data Fig. 1a for the adopted reference frame) and is used as a wavefront division beamsplitter[20,33,34]. The two split beams propagating downstream of M1 are recombined on the sample by mirrors M2 and M3. The design position of the mirrors corresponds to a parallelogram geometry, where the splitting angle after M1 (2α) equals the crossing angle at the sample (2θ) and the M1–M2 distance (d1) equals the M3–sample distance (d2). In this case the optical path difference (ΔL) between FEL path 1 (FP1) and FEL path 2 (FP2) vanishes, ensuring the time coincidence of the crossed FEL beams. The design values of the system are $2\alpha = 2\theta = 6^\circ$ and d1 = d2 = 125 mm. The system has several motorized degrees of freedom to adjust the mirror positions and angles during the experiment: (1) the pitch and roll angles of all mirrors can be independently controlled in the $\pm 3.5^\circ$ range with a resolution of ~ 50 μrad by piezo-electric steering motors; (2) the y-position of M1 can be changed in the ± 20 mm range with ~ 2 μm resolution; (3) the x and y positions of M2 and M3 can be independently varied in the ± 9.5 mm range by linear piezomotors with a resolution of ~ 0.2 μm; (4) each mirror can be removed from the FEL paths; (5) the whole system can be adjusted in x, y and z. Furthermore, the sample is mounted on a (x, y, z, pitch, roll) adjustment stage, which also allows the SiO_2 sample to be replaced with a fluorescent screen, a pinhole (PHsam) or a Si_3N_4 reference sample (for cross-correlation measurements described further below). Such motors can be used to change the value of 2θ (in the 3°–9° range) keeping fixed the arrival time difference of the two FEL pulses ($\Delta t_{\text{EUV-EUV}}$) or vice versa; see sketches in Extended Data Fig. 1d and e. The possibility of varying $\Delta t_{\text{EUV-EUV}}$ (in the ± 0.2 ps range for $2\theta \approx 6^\circ$) at fixed 2θ makes our set-up *de facto* a compact split and delay device, with the advantage of the angular discrimination of the two FEL pulses.

The system was pre-aligned using a Ti:sapphire laser to simulate the incoming FEL beam and optimizing the second harmonic signal generated by a nonlinear crystal placed at the sample position (this signal is sensitive to the time–space superposition of the beams coming from FP1 and FP2). The optimized geometry was then referenced by a set of 2-mm-diameter pinholes (PHin, PHout, PH1, PH2, PH3, PH4, PHsam) before plugging the set-up into the end-station; PHsam can be inserted in place of the sample, while PHsam and PHout can be independently scanned in y in order to determine both the parallelism and the interaxis distance (h) between the trajectory of the incoming FEL beam (once M2 has been removed from the scattering plane) and FP2 downstream of M3. This set of pinholes provides an accuracy of about 0.2° in the determination of 2θ. Indeed, from atomic force microscopy measurements on permanent gratings (carried out after the experiment) we were able to determine the actual value of 2θ (that is, $6.16^\circ \pm 0.02^\circ$) used in the FWM measurements. Such deviations with respect to the design value can be compensated by changing the FEL wavelength (λ_{EUV})[35] and/or the incidence angle of the optical pulse at the sample (θ_B).

A sketch of the system (external to the end-station) used to control the optical pulse is shown in Extended Data Fig. 1b. This pulse is a fraction of the one that triggers the FEL emission, thus ensuring 'jitter-free' conditions, that is, the FEL–optical timing jitter (<10 fs)[36] is much shorter than the pulse duration. An optical delay line (DL) is used to set the time delay (Δt) between the FEL pulses and the optical pulse in the -10 to $+300$ ps range. The optical beam is frequency doubled at 392.8 nm wavelength by a BBO (barium boron oxide) crystal and routed into the sample by a reflective mirror (ML; grazing angle $\sim 22.5^\circ$) and a focusing lens (FL; focal length 600 mm), whose focal plane contains the crossing point of the two FEL beams. ML can be translated along the y-direction in order to change θ_B (in the 47°–51° range) without varying the position of the laser spot at the sample. This degree of freedom is coupled to Δt, so that a change in θ_B has to be compensated by an adjustment of the DL to keep fixed the actual value of Δt. Bandpass filters (BFs) are used to remove the residual fundamental radiation at 785.6 nm wavelength, while a beamsplitter (BS) provides a reference beam for intensity monitoring and active stabilization[37], the latter is ensured by a tip-tilt stage action on the MLS mirror.

An accurate determination of the temporal and spatial coincidence of the three input beams is obtained by cross-correlation measurements[38]. These are based on FEL-induced transient optical reflectivity changes ($\Delta R/R$) from a Si_3N_4 reference sample. In such measurements we first removed M3 from FP2 to collect the cross-correlation trace between the optical pulse and the FEL pulse coming from FP1. The half drop of the $\Delta R/R$ profile was assumed as $\Delta t = 0$. Afterwards we removed M2 and inserted M3 in order to collect the cross-correlation trace associated with the FEL pulse coming from FP2. Iterative measurements, carried out by varying the pitch of M1 and roto-translating M3 keeping fixed the FEL trajectory downstream of M3 (see Extended Data Fig. 1e), allowed us to set the condition $\Delta t_{\text{EUV-EUV}} = 0$. Furthermore, since the amplitude of the $\Delta R/R$ drop is proportional to the FEL fluence at the sample[38,39], such measurements also allowed us to equalize the fluence level of the two FEL beams in the interaction region. Extended Data Fig. 1f and g reports typical $\Delta R/R$ traces collected out of (or in) time coincidence and without (or with) similar fluence levels in the interaction region. Once we had set the time-space coincidence of the three beams, we adjusted θ_B to optimize the phase matching conditions necessary to observe the FWM signal. The latter was recorded by a Princeton Instrument PI-MTE back illuminated Charge Coupled Device (CCD) camera with frame format 2,048 $\times$ 2,048 pixels and 13.5 $\times$ 13.5 μm^2 pixel size. The detector was positioned ~ 250 mm downstream of the sample and oriented at an angle of $\sim 49^\circ$. A 15 $\times$ 5 mm^2 slit was placed in front of the CCD in order to reduce spurious light, which was mainly coming from diffuse scattering of the optical pulse (transmitted through the sample) that impinged on the wall of the experimental chamber.

AFM measurements on permanent gratings. Extended Data Fig. 2a–c shows atomic force microscopy (AFM) topographies of a 8 $\times$ 8 μm^2 sample area not irradiated by the FEL pulses, irradiated by ~ 300 shots at a fluence larger than 50 mJ cm^{-2}, and continuously irradiated by FEL pulses at low fluence, respectively. AFM scans have been performed with a XE-100 (Park Instruments) instrument, in contact mode using commercial cantilevers (Mikromasch, CSC38, spring constant $k = 0.03$–0.09 N m^{-1}) with samplings of 1,024 pixels in both directions. Extended Data Fig. 2d–f report the corresponding depth profiles of the sample surface. The non-irradiated sample shows a roughness of about 0.6 nm root-mean-square and 5 nm peak-to-valley. A grating with a peak-to-valley amplitude of about 18 nm and a period of 256.7 ± 0.9 nm (corresponding to $2\theta = 6.16^\circ \pm 0.02^\circ$) is clearly visible in the sample surface irradiated at high FEL fluence. The high grating visibility after multi-shot illumination indicates that the shot-by-shot fluctuations in the optical path difference (that is, in the relative phase) between the crossed FEL pulses are lower than λ_{EUV}. Such a low phase jitter might in principle allow for lithographic applications. Indeed, the interference between coherent FEL pulses would permit permanent gratings to be imprinted with a pitch as short as a few nanometres by exploiting larger crossing angles and shorter FEL wavelengths. Extended Data Fig. 2g–i show the power spectral densities (PSD) of the data reported in Fig. 2d–f. No specific frequencies are found in the non-irradiated sample (the peak at ~ 35 μm^{-1} is an artefact due to the electronic noise of the employed AFM device), while the first peak at ~ 0.5 μm^{-1} in the PSD, shown in Extended Data Fig. 2h, reveals a modulation with a period of ~ 2 μm, most probably due to diffraction effects from the mirrors. The sample surface irradiated at low fluence shows a roughness similar to that of the non-irradiated sample, and frequencies ascribable to the formation of permanent gratings are absent.

Data analysis. For each probed Δt-value we acquired 5 CCD images (exposure time: 1 min) with the FEL on and 2 with the FEL off. The FEL off images were used to remove the background due to diffuse scattering of the optical laser, mainly due to the fraction of the laser pulse transmitted through the sample, which then impinged on the wall of the experimental chamber. Such frequent and accurate determination of the background allowed us to account for small drifts in the CCD response during the Δt-scan, which lasted several hours. After background subtraction we applied a low-pass Fourier filter in order to reduce the noise and improve the contrast. In order to convert the ADC counts, recorded by the CCD, into incoming photons we assumed a sensitivity of 8.5 photons per ADC count, which also takes into account the detector quantum efficiency. The total number of photons in the FWM beam was determined by fitting the peak in each image with a two-dimensional Gaussian function plus a flat background and then calculating the volume under the Gaussian surface. The volume corresponding to each image was then averaged taking into account the normalization to the corresponding mean squared FEL intensity; error bars were estimated as one standard deviation of the set of CCD images corresponding to the same Δt value (in a few cases we have a single image and we hence set the error bar equal to zero). At negative time delays ($\Delta t < 0$) and within the signal-to-noise of the employed CCD detector, the 'FEL on – FEL off' images can be fitted only by the background (see also Extended Data Fig. 3). Before starting the time delay scan, shown in Fig. 3a, we optimized the FWM signal at $\Delta t = 0$ by small tweaks of mirror positions and angles. The intensity stability of the FEL was typically better than 7% during the acquisition of the 5 FEL on images and about 20% throughout the whole Δt-scan. A few images were discarded due to occasional machine faults occurring during the acquisitions.

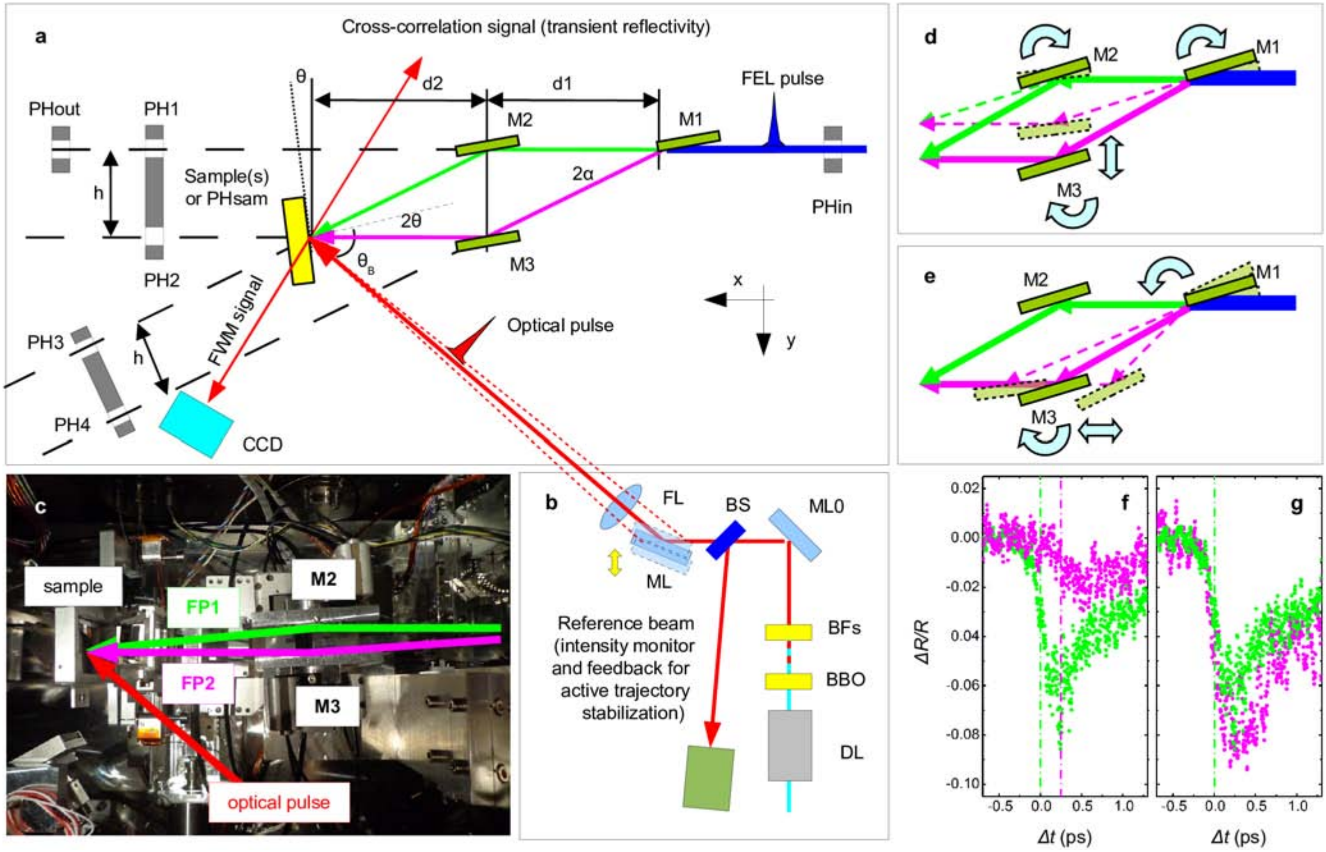
In a proof-of-principle, we assumed that the time dependence of the FWM signal is due to impulsively stimulated vibrational modes, which can be observed in the time-dependent coherent FWM signal[3,40]. We hence modelled the time dependent intensity of the normalized FWM signal ($\alpha I_{\mathrm{FWM}}(\Delta t)/I_{\mathrm{opt}}$; where I_{opt} and α are the intensity of the incoming optical beam and the normalizing factor quoted in the main text, respectively) with the following function:

$$\frac{\alpha I_{\mathrm{FWM}}(\Delta t)}{I_{\mathrm{opt}}} = R(\Delta t)\otimes\left|A_0\delta(\Delta t)+\sum\nolimits_i A_i e^{-\Delta t/\tau_i}\sin(2\pi v_i\Delta t)\right|^2 = A_0^2 R(\Delta t) + R(\Delta t)\otimes\left|\sum\nolimits_i A_i e^{-\Delta t/\tau_i}\sin(2\pi v_i\Delta t)\right|^2 \quad (1)$$

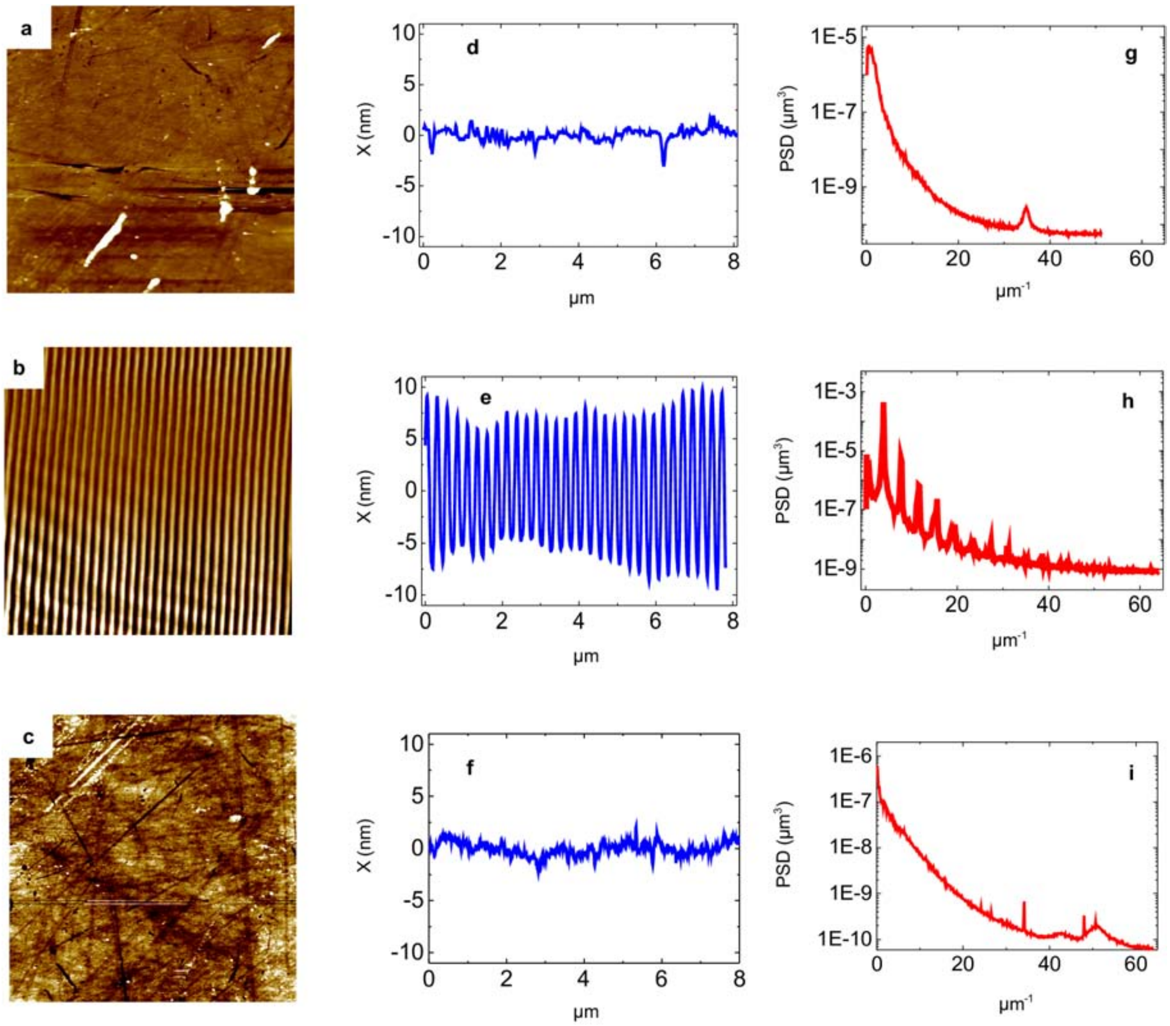
where A_0 and A_i are scaling factors, $\otimes$ is the convolution operator, $R(\Delta t)$ is the instrumental response function, while v_i and τ_i are the frequencies and damping times of the impulsive stimulated modes, respectively. Since an independent determination of $R(\Delta t)$ is not available, it was assumed to be a Gaussian function of unit area with a full-width at half-maximum of 135 fs, which is equal to that of the $\Delta t = 0$ peak in the FWM signal ($R_{cc}(\Delta t)$; see Fig. 3a). Hence, the $A_0\delta(\Delta t)$ term accounts for processes occurring at timescales substantially shorter than $\sim$100 fs. The $e^{-\Delta t/\tau_i}\sin(2\pi v_i\Delta t)$ terms describe the time evolution of the impulsively excited vibrational modes. Considering that the excitable modes are those having vibrational periods (v_i^{-1}) longer than the time duration of the stimulating FEL pulses (60–80 fs)[3,40], the observed FWM signal could be related to: (1) acoustic modes ($v_{LA} = c_s/L \approx 23.2$ GHz, where $c_s = 5{,}970\ \mathrm{m\,s^{-1}}$ and $L = 256.8$ nm are the sound velocity and the transient grating pitch, respectively); (2) Raman modes within the broad band (3–14 THz) associated with tetrahedral bendings[41] and (3) F_1 modes ($\sim$1 THz) involving coupled rotations of SiO_4 tetrahedra[42,43]. Starting from such trial frequencies, we used equation (1) to optimize the parameters that better describe the time structure of the FWM signal, once the $A_0{}^2R(\Delta t)$ term was subtracted. The coarse sampling and the limited Δt-range prevented the direct determination of acoustic parameters, which were fixed to the expected values, that is,: $v_{LA} = 23.2$ GHz and $\tau_{LA} > 1$ ns (we note that in the probed Δt-range $e^{-\Delta t/\tau_{LA}} \approx 1$), while A_{LA} was fixed to 1×10^{-4} in order to fit in the observed magnitude of the FWM signal at $\Delta t > 2$ ps. The optimized values of the other parameters are: $A_1/A_{LA} = 1.4 \pm 0.2$, $v_1 = 1.15 \pm 0.15$ THz, $\tau_1 = 5 \pm 2$ ps, $A_2/A_{LA} = 1.4 \pm 0.2$, $v_2 = 4.1 \pm 0.8$ THz and $\tau_2 = 0.15 \pm 0.06$ ps. The value of v_1 roughly matches the characteristic frequency of F_1 modes[42,43], while v_2 is located on the low frequency side of the afore-mentioned broad Raman band, which corresponds to (highly damped) v_{2b} bending modes[41]. We finally note that a slightly worse fit (but still satisfactory) is achieved without the A_2 term, while the addition of other high frequency terms (such as the $v_{4b,4c}$ bendings)[41] does not significantly improve the fitting results.

Sample size. No statistical methods were used to predetermine sample size.

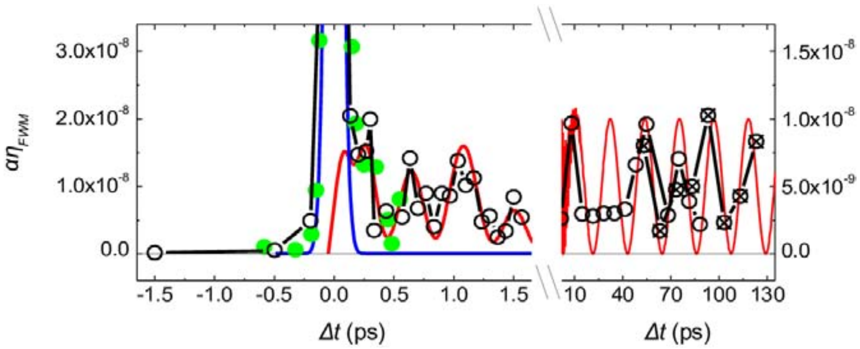
31. Capotondi, F. *et al.* Coherent imaging using seeded free-electron laser pulses with variable polarization: first results and research opportunities. *Rev. Sci. Instrum.* **84,** 051301 (2013).
32. Zangrando, M. *et al.* The photon analysis, delivery, and reduction system at the FERMI@Elettra free electron laser user facility. *Rev. Sci. Instrum.* **80,** 113110 (2009).
33. Cucini, R., Bencivenga, F., Zangrando, M. & Masciovecchio, C. Technical advances of the TIMER project. *Nucl. Instrum. Methods A* **635,** S69–S74 (2011).
34. Cucini, R., Bencivenga, F. & Masciovecchio, C. All-reflective femtosecond optical pump-probe setup for transient grating spectroscopy. *Opt. Lett.* **36,** 1032–1034 (2011).
35. Allaria, A. *et al.* Tunability experiments at the FERMI@Elettra free-electron laser, *New J. Phys.* **14,** 113009 (2012).
36. Danailov, M. B. *et al.* Towards jitter-free pump-probe measurements at seeded free electron laser facilities. *Opt. Express* **22,** 12869–12879 (2014).
37. Cinquegrana, P. *et al.* Optical beam transport to remote location for low jitter pump-probe experiments with free electron laser. *Phys. Rev. ST Accel. Beams* **17,** 040702 (2014).
38. Gahl, C. *et al.* A femtosecond X-ray/optical cross-correlator. *Nature Photon.* **2,** 165–169 (2008).
39. Casolari, F. *et al.* Role of multilayer-like interference effects on the transient optical response of Si_3N_4 films pumped with free-electron laser pulses. *Appl. Phys. Lett.* **104,** 191104 (2014).
40. Ruhman, S., Joly, A. G. & Nelson, K. A. Time resolved observations of coherent molecular vibrational motion and the general occurrence of impulsive stimulated scattering. *J. Chem. Phys.* **86,** 6563–6565 (1987).
41. Kalampounias, A. G., Yannopoulos, S. N. & Papatheodorou, G. N. Temperature-induced structural changes in glassy, supercooled, and molten silica from 77 to 2150 K. *J. Chem. Phys.* **124,** 014504 (2006).
42. Hehlen, B. *et al.* Hyper-Raman scattering observation of the boson peak in vitreous silica. *Phys. Rev. Lett.* **84,** 5355–5358 (2000).
43. Buchenau, U., Nucker, N. & Dianoux, A. J. Neutron scattering study of the low-frequency vibrations in vitreous silica. *Phys. Rev. Lett.* **53,** 2316–2319 (1984).



Extended Data Figure 1 | Experimental set-up for FEL-based FWM measurements. **a**, Top-view layout of the experimental set-up used to split and recombine the FEL beams. **b**, Top-view layout of the experimental set-up used to control the optical beam. **c**, Top-view picture of the set-up: the two FEL paths (FP1 and FP2) downstream of M1 and the trajectory of the optical pulse are indicated. **d**, Sketch of the movements needed to change 2θ keeping $\Delta t_{\mathrm{EUV\text{-}EUV}}$ fixed. **e**, Sketch of the movements needed to change $\Delta t_{\mathrm{EUV\text{-}EUV}}$ keeping 2θ fixed. See Methods for details of **a**–**e**. **f**, **g**, Optical reflectivity changes in Si_3N_4 induced by the FEL beam propagating through FP1 (green dots) and FP2 (magenta dots). In **f**, the mirrors were displaced with respect to the nominal position; a poor time coincidence and a different fluence level in the interaction region can be seen. **g**, Same measurements as in **f** after optimization of the geometry; the superposition of the two traces indicates a large improvement in the time coincidence and a similar FEL fluence in the interaction region.



Extended Data Figure 2 | AFM topographies. AFM topographies of $8 \times 8\,\mu m^2$ areas of the sample surface as follows: in a region that was not irradiated (**a**), in an area irradiated by ~300 FEL shots at a fluence larger than $50\,mJ\,cm^{-2}$ (**b**), and in an area continuously irradiated by FEL pulses at low fluence (**c**). **d**–**f**, Representative depth profiles of the sample surface along the green lines shown in **a**–**c**, respectively. The power spectral densities (PSD) corresponding to data reported in **d**–**f** are shown in **g**–**i**, respectively.



Extended Data Figure 3 | Time sequence of acquired data. Black open and crossed circles connected by lines are data shown in Fig. 3; crossed circles correspond to a scan made several hours after the one corresponding to data shown as open circles; in both scans the time delay was continuously increased. Green dots are data collected before these two scans; here we had not yet optimized the FWM signal at $\Delta t = 0$ (these data are scaled by a factor to fit the peak intensity of the data shown as black circles). Blue and red lines are the same as shown in Fig. 3.

doi:10.1038/nature14342

Measurement of the first ionization potential of lawrencium, element 103

T. K. Sato[1], M. Asai[1], A. Borschevsky[2,3], T. Stora[4], N. Sato[1], Y. Kaneya[1,5], K. Tsukada[1], Ch. E. Düllmann[3,6,7], K. Eberhardt[3,7], E. Eliav[8], S. Ichikawa[1,9], U. Kaldor[8], J. V. Kratz[7], S. Miyashita[10], Y. Nagame[1,5], K. Ooe[11], A. Osa[1], D. Renisch[7], J. Runke[6], M. Schädel[1], P. Thörle-Pospiech[3,7], A. Toyoshima[1] & N. Trautmann[7]

The chemical properties of an element are primarily governed by the configuration of electrons in the valence shell. Relativistic effects influence the electronic structure of heavy elements in the sixth row of the periodic table, and these effects increase dramatically in the seventh row—including the actinides—even affecting ground-state configurations[1,2]. Atomic *s* and $p_{1/2}$ orbitals are stabilized by relativistic effects, whereas $p_{3/2}$, *d* and *f* orbitals are destabilized, so that ground-state configurations of heavy elements may differ from those of lighter elements in the same group. The first ionization potential (IP_1) is a measure of the energy required to remove one valence electron from a neutral atom, and is an atomic property that reflects the outermost electronic configuration. Precise and accurate experimental determination of IP_1 gives information on the binding energy of valence electrons, and also, therefore, on the degree of relativistic stabilization. However, such measurements are hampered by the difficulty in obtaining the heaviest elements on scales of more than one atom at a time[3–5]. Here we report that the experimentally obtained IP_1 of the heaviest actinide, lawrencium (Lr, atomic number 103), is $4.96^{+0.08}_{-0.07}$ electronvolts. The IP_1 of Lr was measured with ^{256}Lr (half-life 27 seconds) using an efficient surface ion-source and a radioisotope detection system coupled to a mass separator. The measured IP_1 is in excellent agreement with the value of 4.963(15) electronvolts predicted here by state-of-the-art relativistic calculations. The present work provides a reliable benchmark for theoretical calculations and also opens the way for IP_1 measurements of superheavy elements (that is, transactinides) on an atom-at-a-time scale.

The chemical characterization of heavy elements at the end of the actinides and of all transactinides, performed to probe their positions in the periodic table[4–10], has been, so far, conducted by rapid chemical separation techniques, such as gas-phase and liquid-phase chromatography. The influence of relativistic effects on electronic orbitals has been inferred indirectly, through a comparison of the chemical properties of the heavy elements with those of their lighter homologues and those predicted by theoretical calculations. The first ionization potential, one of the most fundamental physical and chemical properties of an element, gives direct information about the binding energy of an electron in the outermost electronic orbital of an atom. Accurate IP_1 values of heavy elements provide crucial tests for our understanding of their electronic structure. IP_1 values of weighable amounts of nuclear-reactor-produced heavy elements up to einsteinium (atomic number $Z = 99$) have been successfully measured by resonance ionization mass spectroscopy (RIMS)[11,12]. RIMS was also used in an investigation of fermium (Fm, $Z = 100$) with a sample of 2.7×10^{10} atoms of ^{255}Fm (half-life $T_{1/2} = 20.1$ h). In that experiment, the atomic level structure, but not the IP_1, was determined[13]. Recently, resonance ionization laser ion source (RILIS) studies optimized for short-lived atoms made it possible to determine the IP_1 of astatine (At, $Z = 85$) using ^{199}At ($T_{1/2} = 7.2$ s) produced in the proton-induced spallation reaction of uranium[14]. IP_1 values of heavy elements with $Z \geq 100$, however, could not be determined experimentally, because production rates drastically decrease for elements as their atomic number increases. The study of these elements therefore requires new techniques, on an atom-at-a-time scale.

The ground-state electronic configuration of Lr is predicted to be $[\mathrm{Rn}]5f^{14}7s^27p_{1/2}$, in contrast to that of its lanthanide homologue Lu, $[\mathrm{Xe}]4f^{14}6s^25d$, as the $7p_{1/2}$ orbital is expected to be stabilized below the $6d$ orbital in Lr by strong relativistic effects[15–19]. The determination of IP_1 sheds light on the important role of relativistic effects in heavy elements by comparison with theoretical predictions. For Lr, theory predicts an exceptionally low IP_1 value. A sufficiently long-lived and detectable isotope for ionization experiments is ^{256}Lr ($T_{1/2} = 27$ s). It is produced at a rate of one atom per several seconds in the fusion-evaporation reaction of a ^{249}Cf target with a ^{11}B beam[20]. With this constraint, a new and highly efficient experimental set-up based on the ionization and detection of the $^{256}Lr^+$ ion has been devised and implemented to determine the IP_1 value of Lr.

The surface ionization process takes place on a solid surface kept at high temperature, which is coupled to an on-line mass separator; that is, an atom is ionized to the 1+ charge state via the interaction with a solid (metal) surface at high temperature and is selectively mass-separated from nuclear reaction by-products. Figure 1 depicts the experimental set-up schematically. Short-lived ^{256}Lr atoms recoiling off a ^{249}Cf target were promptly transported to the ionization site (ionization cavity) by He/CdI_2 gas-jet transport, and surface-ionized ^{256}Lr atoms were extracted and mass-separated. The number of ^{256}Lr ions after the mass-separation was determined by α-particle spectroscopy. Experimental details are provided in the Methods section and in ref. 20.

Based on the Saha-Langmuir equation[21,22], an analytical model[23] describes the surface ionization process in a hollow tube (cavity)-type ion-source. The ionization efficiency I_{eff} can be expressed as:

$$I_{\mathrm{eff}} = \frac{N\exp\left(\dfrac{\phi - \mathrm{IP}_1^*}{kT}\right)}{1 + N\exp\left(\dfrac{\phi - \mathrm{IP}_1^*}{kT}\right)} \tag{1}$$

where ϕ is the work function, which is material-dependent, k the Boltzmann constant, T the temperature of the ionizing surface, and N a parameter that depends on the effective number of atom–surface interactions in the cavity. IP_1*, the effective IP_1, is directly related to IP_1 as[22]:

$$\mathrm{IP}_1^* = \mathrm{IP}_1 - kT\ln\left(\frac{Q_{\mathrm{i}}}{Q_0}\right) \tag{2}$$

where Q_{i} and Q_0 are the partition functions at a given temperature for the ion and the atom, respectively, which can be calculated using excitation

[1]Japan Atomic Energy Agency (JAEA), Tokai, Ibaraki 319-1195, Japan. [2]Centre for Theoretical Chemistry and Physics, New Zealand Institute for Advanced Study, Massey University, 0745 North Shore MSC, Auckland, New Zealand. [3]Helmholtz-Institut Mainz, 55099 Mainz, Germany. [4]ISOLDE, CERN, CH-1211 Geneva 23, Switzerland. [5]Graduate School of Science and Engineering, Ibaraki University, Mito, Ibaraki 310-8512, Japan. [6]GSI Helmholtzzentrum für Schwerionenforschung, 64291 Darmstadt, Germany. [7]Institut für Kernchemie, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany. [8]School of Chemistry, Tel Aviv University, 69978 Tel Aviv, Israel. [9]Nishina Center for Accelerator-Based Science, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. [10]Graduate School of Science, Hiroshima University, Kagamiyama, Higashi-Hiroshima 739-8526, Japan. [11]Institute of Science and Technology, Niigata University, Niigata 910-2181, Japan.

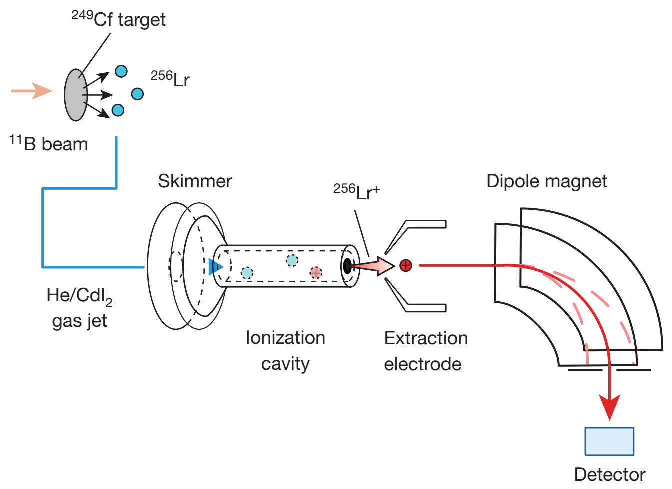


Figure 1 | Schematic experimental set-up used to measure the IP$_1$ of Lr on an atom-at-a-time scale. ^{256}Lr atoms produced in the nuclear reaction ^{249}Cf(^{11}B, 4n) are transported (blue line) by the He/CdI$_2$ gas jet to a tantalum ionization cavity via a skimmer component. The ^{256}Lr atoms are ionized on the tantalum surface. The ^{256}Lr$^+$ ions produced are extracted electrostatically from the cavity by the extraction electrode, accelerated by 30 kV, and mass-separated by a dipole magnet. See Methods for details. The number of mass-separated ^{256}Lr$^+$ ions is determined by α-particle spectroscopy at the detector position.

energies and statistical weights of their ground and excited states. Tantalum (Ta) was chosen as the cavity material. The ionization experiments were conducted at $T = 2{,}700$ K and 2,800 K. For ^{256}Lr, I_{eff} values of (33 ± 4)% and (36 ± 7)%, respectively, were determined by the procedure given in ref. 20.

The following procedure was applied to determine the value of the free parameter N in equation (1): short-lived lanthanide and alkali isotopes ^{142,143}Eu, ^{143}Sm, ^{148}Tb, ^{153,154}Ho, ^{157}Er, ^{162}Tm, ^{165}Yb, ^{168}Lu and ^{80}Rb were produced in nuclear reactions of ^{11}B beams with target materials of ^{136}Ce, ^{141}Pr, ^{142}Nd, ^{147}Sm, Eu, ^{156}Gd, ^{159}Tb, ^{162}Dy and Ge, respectively, and their I_{eff} values were experimentally determined at $T = 2{,}700$ K and 2,800 K. Figure 2 shows the I_{eff} values at 2,700 K as a function of IP$_1$*. The IP$_1$* value for each element was calculated with equation (2). Energies and statistical weights of low-lying states in the ion and the atom of each element were taken from the National Institute of Standards and Technology (NIST) atomic database[24]. The I_{eff} values determined for all isotopes were best fitted with equation (1) using N values of 43 ± 3 and 50 ± 3 at $T = 2{,}700$ K and 2,800 K, respectively.

The Lr IP$_1$* values of $5.29^{+0.08}_{-0.07}$ eV and $5.33^{+0.11}_{-0.10}$ eV were determined from equation (1) at $T = 2{,}700$ K and 2,800 K, respectively. The result at 2,700 K is illustrated in Fig. 2. Errors on the IP$_1$* values mainly came from three sources of uncertainty: surface temperatures, I_{eff} (which is based on counting statistics) and fitting procedures. The Lr IP$_1$ can be calculated from IP$_1$* using equation (2) with Q_i and Q_0. No experimental data on excited states of the Lr atom and ion are available. Thus, the energies and statistical weights for calculating Q_i and Q_0 were taken from relativistic Fock space coupled cluster (FSCC) calculations[18]. The average absolute error for the 20 lowest excitation energies of Lu (where comparison with experiment is possible) was 0.05 eV using the same approach[18]. We expected a similar accuracy for the predicted transition energies of Lr. The evaluated values of $kT\ln(Q_i/Q_0)$ for Lr at $T = 2{,}700$ K and 2,800 K are $-0.34^{+0.06}_{-0.04}$ and $-0.36^{+0.06}_{-0.04}$, respectively. The errors include uncertainties in the calculated excitation energies indicated in ref. 18, 0.087 eV (700 cm^{-1}) for each state, and in the temperatures. From this, IP$_1$ values of $4.95^{+0.10}_{-0.08}$ eV and $4.97^{+0.13}_{-0.11}$ eV were obtained at $T = 2{,}700$ K and 2,800 K, respectively. Based on these results, our experimentally determined value for IP$_1$ of Lr is $4.96^{+0.08}_{-0.07}$ eV.

A theoretical calculation of the IP$_1$ of Lr was also performed, using the relativistic coupled cluster approach with single, double, and perturbative triple excitations (DC CCSD(T)), and corrected for the Breit contribution

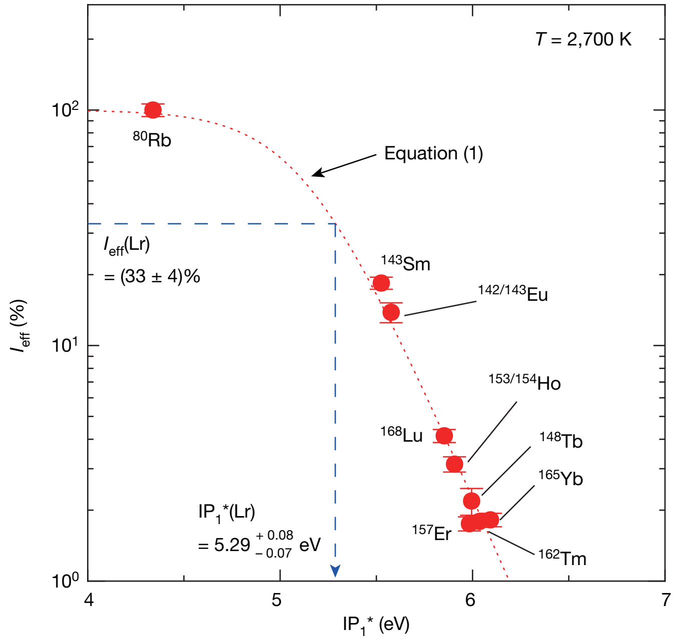


Figure 2 | The ionization efficiency (I_{eff}) of various short-lived isotopes as a function of the effective IP$_1$ (IP$_1$*) at 2,700 K. The short-dashed curve is obtained by fitting the experimental data using equation (1), which describes a relationship between I_{eff} and IP$_1$* in our present system. The position of the measured I_{eff} value of Lr, (33 ± 4)%, is also shown. From the fitted equation (1) with $N = 43 \pm 3$, IP$_1$* of Lr is calculated to be $5.29^{+0.08}_{-0.07}$ eV. This corresponds to an IP$_1$ value of $4.95^{+0.10}_{-0.08}$ eV at 2,700 K. Error bars, ±1s.d.

and Lamb shift (Methods). The calculated $7s^27p_{1/2}$ level of Lr is lower by ~180 meV than $7s^26d_{3/2}$, confirming earlier identification of the former as the atomic ground state. This is due to relativistic effects; a non-relativistic calculation puts the energy of the $7s^26d$ configuration about 2.2 eV below that of $7s^27p$. To assess the accuracy of our predicted IP$_1$ of Lr, the same approach was applied to its lighter homologue Lu. The calculated ground state of the latter is experimentally confirmed as $6s^25d$. The Lr$^+$ ion has a closed-shell [Rn]$5f^{14}7s^2$ ground-state configuration. The IP$_1$ was obtained by taking the difference between the calculated energies of the neutral state and the 1+ state. The calculated IP$_1$ values are 5.418 eV for Lu and 4.963 eV for Lr. Corrections for the Breit contribution were 6 meV for Lu and −12 meV for Lr, and corrections for the Lamb shift were 0.3 meV for Lu and 16 meV for Lr (ref. 25). The calculated IP$_1$ for Lu is in very good agreement with the experimental IP$_1$ of 5.425871(12) eV (ref. 26); similar accuracy is expected for the calculated IP$_1$ of Lr.

The experimental and calculated IP$_1$ results obtained in our work are shown in Table 1 together with earlier theoretical predictions. It should be noted that the calculated excitation energies of Lr, which we used to

Table 1 | Theoretical and experimental IP$_1$ values of Lr

Reference	Year	IP$_1$ (eV)	Method
Ref. 16	1995	4.887	DCB+FSCC*
Ref. 27	1998	5.28	RECP+CASSCF+ACPF+Δ_{SO}†
Ref. 28	2003	4.80	RECP+CASSCF+ACPF+Δ_{SO}†
Ref. 18	2007	4.893	DCB+FSCC*
Ref. 19	2014	4.934	CI+all-order‡
This work		4.963(15)	DC CCSD(T)+Breit+Lamb shift§
This work		$4.96^{+0.08}_{-0.07}$	Experimental

*Dirac Coulomb Breit (DCB) Hamiltonian combined with the Fock space coupled cluster (FSCC) approach.
†Relativistic effective core potentials (RECP) combined with complete active space multiconfiguration self-consistent field (CASSCF) calculations with subsequent multi-reference averaged coupled-pair functional (ACPF) calculations and spin orbit corrections (Δ_{SO}).
‡Configuration interaction (CI) method combined with the linearized single-double coupled cluster method (all-order).
§Dirac Coulomb (DC) Hamiltonian combined with CCSD(T) approach and corrected for the Breit contribution and the Lamb shift.

get Q_i and Q_0 values to derive the experimental IP_1 from IP_1*, were obtained with a method different to the one employed here for the calculation of the IP_1 itself. As the two calculations are independent, we can compare the present experimental and theoretical IP_1 values. Our experimental result on the first ionization potential of Lr of $4.96^{+0.08}_{-0.07}$ eV is in excellent agreement with the theoretical value of 4.963(15) eV also obtained in this work.

Thus, we have experimentally shown that the IP_1 of Lr is distinctly lower than that of Lu. Lr, the heaviest actinide, has the lowest IP_1 value of all lanthanides and actinides; this quantitatively reflects and confirms the theoretically predicted situation of closed $5f^{14}$ and $7s^2$ shells with an additional weakly-bound electron in the valence orbital. We note that the surface ionization method, successfully applied here to determine the IP_1 of Lr, can provide experimental data that can benchmark quantum chemical calculations of the heaviest elements. In addition, it opens up new perspectives on determining basic atomic properties of the superheavy elements.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 November 2014; accepted 6 February 2015.

1. Pyykkö, P. Relativistic effects in structural chemistry. *Chem. Rev.* **88,** 563–594 (1988).
2. Pershina, V. in *Relativistic Methods for Chemists* (eds Barysz, M. & Ishikawa, Y.) 451–520 (Challenges and Advances in Computational Chemistry and Physics 10, Springer, 2010).
3. Oganessian, Yu. Heaviest nuclei from ^{48}Ca-induced reactions. *J. Phys. G* **34,** R165–R242 (2007).
4. Schädel, M. & Shaughnessy, D. (eds) *The Chemistry of Superheavy Elements* 2nd edn (Springer, 2014).
5. Türler, A. & Pershina, V. Advances in the production and chemistry of the heaviest elements. *Chem. Rev.* **113,** 1237–1312 (2013).
6. Schädel, M. *et al.* Chemical properties of element 106 (seaborgium). *Nature* **388,** 55–57 (1997).
7. Eichler, R. *et al.* Chemical characterization of bohrium (element 107). *Nature* **407,** 63–65 (2000).
8. Düllmann, Ch. E. *et al.* Chemical investigation of hassium (element 108). *Nature* **418,** 859–862 (2002).
9. Eichler, R. *et al.* Chemical characterization of element 112. *Nature* **447,** 72–75 (2007).
10. Even, J. *et al.* Synthesis and detection of a seaborgium carbonyl complex. *Science* **345,** 1491–1493 (2014).
11. Köhler, S. *et al.* First experimental determination of the ionization potentials of berkelium and californium. *Angew. Chem. Int. Edn Engl.* **35,** 2856–2858 (1996).
12. Peterson, J. *et al.* Determination of the first ionization potential of einsteinium by resonance ionization mass spectroscopy (RIMS). *J. Alloy. Comp.* **271–273,** 876–878 (1998).
13. Sewtz, M. *et al.* First observation of atomic levels for the element fermium (Z = 100). *Phys. Rev. Lett.* **90,** 163002 (2003).
14. Rothe, S. *et al.* Measurement of the first ionization potential of astatine by laser ionization spectroscopy. *Nature Commun.* **4,** 1835 (2013).
15. Desclaux, J.-P. & Fricke, B. Relativistic prediction of the ground state of atomic lawrencium. *J. Phys.* **41,** 943–946 (1980).
16. Eliav, E., Kaldor, U. & Ishikawa, Y. Transition energies of ytterbium, lutetium, and lawrencium by the relativistic coupled-cluster method. *Phys. Rev. A* **52,** 291–296 (1995).
17. Zou, Y. & Fischer, C. F. Resonance transition energies and oscillator strengths in lutetium and lawrencium. *Phys. Rev. Lett.* **88,** 183001 (2002).
18. Borschevsky, A. *et al.* Transition energies of atomic lawrencium. *Eur. Phys. J. D* **45,** 115–119 (2007).
19. Dzuba, V. A., Safronova, M. S. & Safronova, U. I. Atomic properties of superheavy elements No, Lr, and Rf. *Phys. Rev. A* **90,** 012504 (2014).
20. Sato, T. K. *et al.* First successful ionization of Lr (Z = 103) by a surface-ionization technique. *Rev. Sci. Instrum.* **84,** 023304(5) (2013).
21. Stora, T. *Radioactive Ion Sources*. CERN-2013-007, 331–349 http://cds.cern.ch/record/1693046 (accessed November 2014).
22. Zandberg, É. Y. & Ionov, N. I. Surface ionization. *Sov. Phys. Usp.* **2,** 255–281 (1959).
23. Kirchner, R. On the thermoionization in hot cavities. *Nucl. Instrum. Methods A* **292,** 203–208 (1990).
24. Kramida, A., Ralchenko, Y., Reader, J. & Team, N. A. *NIST Atomic Spectra Database* version 5.1 (2013); http://physics.nist.gov/asd (accessed July 2014).
25. Shabaev, V. M., Tupitsyn, I. I. & Yerokhin, V. A. QEDMOD: Fortran program for calculating the model Lamb-shift operator. *Comput. Phys. Commun.* **189,** 175–181 (2015).
26. Maeda, H., Mizugai, Y., Matsumoto, Y., Suzuki, A. & Takami, M. Highly excited even Rydberg series of Lu I studied by two-step laser photoionization spectroscopy. *J. Phys. B* **22,** L511–L516 (1989).
27. Liu, W., Küchle, W. & Dolg, M. *Ab initio* pseudopotential and density-functional all-electron study of ionization and excitation energies of actinide atoms. *Phys. Rev. A* **58,** 1103–1110 (1998).
28. Cao, X., Dolg, M. & Stoll, H. Valence basis sets for relativistic energy-consistent small-core actinide pseudopotentials. *J. Chem. Phys.* **118,** 487–496 (2003).

Acknowledgements We thank the JAEA tandem accelerator crew for supplying intense and stable beams for the experiments. The ^{249}Cf was made available by H. Nitsche (Univ. California, Berkeley); it was produced in the form of ^{249}Bk through the former Transplutonium Element Production Program at Oak Ridge National Laboratory (ORNL) under the auspices of the Director, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division of the US Department of Energy. Financial support by the Helmholtz-Institut Mainz is acknowledged. This work has been partly supported by the Grant-in-Aid for Scientific Research (C) no. 26390119 of the Ministry of Education, Science, Sports and Culture (MEXT).

Author Contributions T.K.S., M.A., Y.N. and M.S. prepared the main part of the manuscript, A.B., E. E. and U.K. contributed to the theory part, and T.S. to the experimental part. C.E.D. and J.V.K. commented on the manuscript. T.K.S, M.A., T.S., N.S., K.T. and S.I. developed the surface ion-source in the ISOL setup at the JAEA tandem accelerator facility. T.K.S. and M.A. were responsible for data acquisition and analysis. T.S. commented on ion-source optimizations and the data analysis procedure. K.T. prepared the ^{249}Cf target. K.E., J.R., P.T.-P., C.E.D. and N.T. separated and provided the ^{249}Cf for the target. The on-line experiments were performed by T.K.S., M.A., N.S., Y.K., K.T., S. I., S.M., Y.N., K.O., A.O., D.R., M.S. and A.T., while theoretical calculations were carried out by A.B., E.E. and U.K. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.N. (nagame.yuichiro@jaea.go.jp).

METHODS

Ionization experiments. The set-up consists of a target-recoil chamber coupled to an aerosol gas-jet transport system, a surface ion-source, a mass separator, and a detection system for nuclear decays[20]. For the Lr experiment, a ^{249}Cf target (thickness 260 μg cm^{-2}) in the target-recoil chamber was irradiated with a 67.9-MeV $^{11}B^{4+}$ beam delivered from the Tandem accelerator at the Japan Atomic Energy Agency (JAEA), Tokai[20,29]. ^{256}Lr atoms, recoiling from the target, attached onto CdI_2 particles produced by sublimation of CdI_2, were transported to the ionization cavity of the ion-source installed in the Isotope Separator On-Line (JAEA-ISOL)[20,30]. Before entering the ionization cavity, aerosol particles with the attached ^{256}Lr passed through a skimmer structure, installed to remove the He carrier gas to achieve high vacuum conditions at the ion-source (typically 2×10^{-2} Pa). In the cavity, the aerosol particles were vaporized and ^{256}Lr atoms were surface ionized. The temperature of the cavity was monitored with a calibrated radiation thermometer with ±50 K accuracy. The cavity can be heated up to 2,850 K. Ions were extracted and accelerated by an electrostatic potential of 30 kV. $^{256}Lr^{+}$ ions were mass separated in the dipole magnet of the JAEA-ISOL and were transported to the detection device. The nuclear decay of ^{256}Lr was measured with eight pairs of Si PIN photodiodes of a rotating catcher wheel apparatus, MANON (Measurement system of Alpha particle and spontaneous fissioN events ON-line) for efficient α-particle measurements[20].

For the short-lived isotope experiments to obtain a relationship between I_{eff} and IP_1* in the present system, various isotopes, ^{142,143}Eu, ^{143}Sm, ^{148}Tb, ^{153,154}Ho, ^{157}Er, ^{162}Tm, ^{165}Yb, ^{168}Lu and ^{80}Rb, were employed. These isotopes were produced in reactions of ^{11}B with targets consisting of ^{136}Ce, ^{141}Pr, ^{142}Nd, ^{147}Sm, Eu, ^{156}Gd, ^{159}Tb, ^{162}Dy and Ge, respectively. Mass-separated ions were collected on an aluminized Mylar tape in a tape transport system installed between the end of the ISOL beam line and the MANON set-up[20]. A high-purity germanium (HP-Ge) detector was placed at the tape transport system to determine the number of the ions by γ spectroscopy.

Calculation of I_{eff} and IP_1*. To calculate the I_{eff} value of each isotope, the number of atoms transported to the ion-source was determined by a direct-catch measurement. Nuclear reaction products were directly collected at a separate collection site where their radioactivity was measured before the ISOL experiments. Here, the measurement of ^{256}Lr was performed using another MANON set-up. For the measurements of the other isotopes, aerosol particles from the target recoil chamber were collected on a glass fibre filter. The γ-rays of each isotope on the glass fibre filter were measured using a HP-Ge detector. Because of the low IP_1 of Rb, it is justified to assume that the I_{eff} of ^{80}Rb should be 100% in the surface ionization. By using ^{80}Rb as a reference material, I_{eff} of the other elements were calculated[20].

For a calculation of IP_1*, equation (2) is applied to each element. Q_i and Q_0 in equation (2) are described as follows[22],

$$\begin{cases} Q_i = \sum_j g_i^j \exp\left(-\frac{E_i^j}{kT}\right) \\ Q_0 = \sum_j g_0^j \exp\left(-\frac{E_0^j}{kT}\right) \end{cases} \quad (3)$$

where g_i^j and g_0^j are the statistical weights of the jth quantum state of the ion and the atom, respectively. The $j = 0$ corresponds to the ground state. Here, E_i^j and E_0^j are the energies of excitation to the jth quantum state of the ion and the atom, respectively. Excited states whose excitation energy is much higher than kT can be neglected because $\exp\left(-\frac{E}{kT}\right)$ is approaching zero. Using equations (3), Q_i and Q_0 values were calculated. Applying these results to the IP_1* value for each element, the IP_1 value was calculated by equation (2).

Theoretical calculation. The CCSD(T) calculations of the IP_1 of Lu and Lr were carried out using the DIRAC13 computational program package[31], in the framework of the relativistic Dirac-Coulomb Hamiltonian. The IP_1 was obtained as the difference between the calculated energies of the neutral and mono-cation species. The Faegri dual family basis sets of uncontracted Gaussian-type orbitals[32] were used. These sets were augmented by high angular momentum and diffuse functions up to convergence of the calculated IP_1 values. The final basis sets consist of 25*s* 23*p* 15*d* 14*f* 6*g* 3*h* orbitals for Lu and 27*s* 25*p* 17*d* 14*f* 6*g* 3*h* 2*i* orbitals for Lr. The finite nucleus model with Gaussian charge distribution was used[33].

Effects of higher-order terms of the Hamiltonian on the IP_1 values, namely the frequency independent Breit operator and the Lamb shift, were also calculated. These terms are not implemented in the DIRAC13 program; the Breit term was therefore calculated using the TRAFS-3C (Tel-Aviv Relativistic Atomic Fock Space coupled cluster code) package[34], and the Lamb energy shift was obtained by the recently developed effective potential method, implemented in the QEDMOD program[25]. The contributions of the Breit term were 6 meV for Lu and −12 meV for Lr; the effect of the Lamb shift on the IP was −0.3 meV for Lu and 16 meV for Lr. These contributions were added to the CCSD(T) results obtained using the DIRAC13 program.

Sample size. No statistical methods were used to predetermine sample size.

29. Sato, N. *et al.* Production of ^{256}Lr in the ^{249,250,251}Cf + ^{11}B, ^{243}Am + ^{18}O, and ^{248}Cm + ^{14}N reactions. *Radiochim. Acta* **102,** 211–219 (2014).
30. Sato, T. K. *et al.* Development of a He/CdI_2 gas-jet system coupled to a surface-ionization type ion-source in JAEA-ISOL: towards determination of the first ionization potential of Lr (Z = 103). *J. Radioanal. Nucl. Chem.* **303,** 1253–1257 (2015).
31. Visscher, L. *et al.* DIRAC, a relativistic *ab initio* electronic structure program. Release DIRAC13 http://www.diracprogram.org (2013).
32. Faegri, K. Relativistic Gaussian basis sets for the elements K-Uuo. *Theor. Chem. Acc.* **105,** 252–258 (2001).
33. Visscher, L. & Dyall, K. G. Dirac–Fock atomic electronic structure calculations using different nuclear charge distributions. *At. Data Nucl. Data Tables* **67,** 207–224 (1997).
34. Eliav, E., Kaldor, U. & Ishikawa, Y. Open-shell relativistic coupled-cluster method with Dirac-Fock-Breit wave functions: energies of the gold atom and its cation. *Phys. Rev. A* **49,** 1724–1729 (1994).

doi:10.1038/nature14333

A primordial origin for the compositional similarity between the Earth and the Moon

Alessandra Mastrobuono-Battisti[1], Hagai B. Perets[1] & Sean N. Raymond[2,3]

Most of the properties of the Earth–Moon system can be explained by a collision between a planetary embryo (giant impactor) and the growing Earth late in the accretion process[1–3]. Simulations show that most of the material that eventually aggregates to form the Moon originates from the impactor[1,4,5]. However, analysis of the terrestrial and lunar isotopic compositions show them to be highly similar[6–11]. In contrast, the compositions of other Solar System bodies are significantly different from those of the Earth and Moon[12–14], suggesting that different Solar System bodies have distinct compositions. This challenges the giant impact scenario, because the Moon-forming impactor must then also be thought to have a composition different from that of the proto-Earth. Here we track the feeding zones of growing planets in a suite of simulations of planetary accretion[15], to measure the composition of Moon-forming impactors. We find that different planets formed in the same simulation have distinct compositions, but the compositions of giant impactors are statistically more similar to the planets they impact. A large fraction of planet–impactor pairs have almost identical compositions. Thus, the similarity in composition between the Earth and Moon could be a natural consequence of a late giant impact.

Successful models for Moon formation typically require a relatively low-velocity, oblique impact[1] between the proto-Earth and a planetary embryo of up to a few tenths of an Earth-mass ($M_{\oplus}$). Such Moon-forming impacts typically occur at the late stages of planetesimal accretion by the terrestrial planets[2,3]. A circum-terrestrial debris disk is formed from material ejected during these impacts. The composition of the disk, in a typical impact, is dominated by material from the impactor mantle (>60 weight per cent[1,4,5]) with a smaller contribution (typically about 20%) from the proto-Earth. More material can be extracted from the proto-Earth when a slightly sub-Mars-sized body hits a fast-spinning planet that is later slowed down by resonances[16]. The spin should be close to the break-up velocity. Another possible channel producing such mixing of material from both the planet and impactor is the rare collision between two embryos of comparable mass, in which both masses are about half of the Earth's mass[17]. Although these new models can potentially solve some of the composition issues raised by the giant-impact scenario, they do require *ad hoc* assumptions and pose several difficulties (see ref. 3 for a discussion). Here, we focus on typical giant-impact events, in which the Moon aggregates mostly from material originating from the impactor.

Lunar meteorites and rock samples returned by the Apollo mission have a very similar composition to that of the Earth's mantle, across a variety of different isotopes[6–11]. Combining these with the giant-impact simulation results, one can infer that the Moon-forming impactor and the Earth should have had a similar composition. This poses a fundamental difficulty with the giant-impact model for the origin of the Moon, since analysis of material from other Solar System bodies have shown them to differ significantly (>20σ) in composition from that of the Earth (see refs 12 and 13, and ref. 14 for a review). This would suggest that the composition of the Moon-forming impactor should similarly have differed from that of the Earth, in contrast with the giant-impact basic prediction.

Here, we analyse the results of extensive *N*-body simulations of terrestrial planet formation to show that the Earth–Moon composition challenge can be addressed. In particular, we show that the compositions of a large fraction, 20% to 40%, of giant impactors are consistent with being similar to that of the planets they impact. More generally, late giant impactors have compositions more similar to those of the planets they impact than do other planets in the same system, showing large differences (the probability of these two distributions of coming from the same parent distribution is between 1.3×10^{-9} and 6.7×10^{-8}).

To study the compositions of planets and their impactors we analysed 40 dynamical simulations (from ref. 15, using the Mercury code[18]) of the late stages of planetary accretion, following the formation of Jupiter and Saturn, and after all the gas in the protoplanetary disk has been dissipated and/or accreted to the gas giants (see ref. 19 for a recent review). Each simulation started from a disk of 85–90 planetary embryos and 1,000–2,000 planetesimals extending from 0.5 astronomical units (1 AU is the Earth–Sun distance) to 4.5 AU. Jupiter and Saturn are fully formed and have different orbits and inclinations in different sets of simulations (detailed descriptions can be found in ref. 15 and in the Methods). Within 100–200 million years, each simulation typically produced 3–4 rocky planets formed from collisions between embryos and planetesimals. Each of these planets accreted a large number of planetesimals during its evolution. All collisions were recorded and provide a map of each planet's feeding zone (see also ref. 20). Assuming that the initial composition of material in the protoplanetary disk is a function of its position in the initial protoplanetary disk, one can compare the compositions of different bodies formed and evolved in the simulations.

Previous studies explored the compositions of the different planets formed in similar simulations. However, the compositions of impactors on formed planets have hardly been explored. Pahlevan and Stevenson[21] analysed a single statistically limited simulation, which included a total of about 150 particles, and compared the compositions of any impactors on any planets during the simulation (not only giant impacts, due to small number statistics) to the compositions of the planets. They[21] concluded that the scatter among the compositions of the various impactors is comparable to the observed differences between the planets. In particular, they found that none of the planetary impactors in the simulation they analysed had an isotopic composition similar enough to the final planet to yield a composition similarity such as that of the Earth and the Moon.

Using the data from our large set of high-resolution simulations, we compare the composition of each surviving planet with that of its last giant impactor, that is, the last planetary embryo that impacted the planet (typical impactor-to-planet mass ratio in the range 0.2–0.5; see Table 1). We include only the 20 cases where both the impactor and planet are composed of at least 50 particles each, so as to have sufficient statistics. Analysis of the additional data for impactors composed of a smaller number of particles (and hence having smaller statistics for the specific composition) are consistent with the higher-resolution cases

[1]Department of Physics, Technion, Israel Institute of Technology, Haifa 32000, Israel. [2]CNRS, Laboratoire d'Astrophysique de Bordeaux, UMR 5804, F-33270 Floirac, France. [3]Université Bordeaux, Laboratoire d'Astrophysique de Bordeaux, UMR 5804, F-33270 Floirac, France.

Table 1 | The modelled planet–impactor systems and observations of Solar System bodies

Model	Number	M_P ($M_\oplus$)	M_I ($M_\oplus$)	M_I/M_P	N_P	N_I	t_{coll} (Myr)	$C_{cal} \times \Delta^{17}$O (p.p.m.)			Kolmogorov–Smirnov probability		
								0%	20%	40%	0%	20%	40%
cjs15	1	0.94	0.43	0.46	123	97	50.7	13 ± 14	10 ± 13	8 ± 12	0.0039	0.13	0.67
cjs15	2	0.78	0.27	0.35	209	78	80.9	$(-1.05 \pm 0.26) \times 10^2$	-84 ± 23	-63 ± 22	1.3×10^{-7}	3.1×10^{-4}	0.079
cjs1	3	1.25	0.42	0.34	219	73	149.7	64 ± 13	52 ± 11	39.9 ± 9.6	5.3×10^{-11}	8.1×10^{-7}	0.0092
cjs1	4	1.05	0.39	0.37	128	78	186.2	$(-1.97 \pm 0.20) \times 10^2$	$(-1.57 \pm 0.20) \times 10^2$	$(-1.18 \pm 0.20) \times 10^2$	1.1×10^{-29}	5.1×10^{-18}	0.052
cjs1	5*	1.21	0.38	0.31	219	75	123.5	-24 ± 17	-20 ± 15	-15 ± 13	0.0023	0.056	0.19
cjsecc	6	0.94	0.36	0.38	117	79	80.4	-51 ± 34	-40 ± 30	-31 ± 27	3.4×10^{-4}	0.025	0.33
cjsecc	7	1.01	0.32	0.32	144	68	75.5	-12 ± 16	-10 ± 13	-7 ± 12	0.038	0.050	0.20
cjsecc	8	1.02	0.42	0.41	148	89	36.9	13 ± 79	11 ± 71	8 ± 66	0.054	0.21	0.53
eejs15	9	0.70	0.19	0.27	111	52	77.1	9.1 ± 7.4	7.3 ± 6.2	5.5 ± 5.2	0.071	0.18	0.32
eejs15	10	0.55	0.13	0.24	263	65	24.6	98 ± 29	78 ± 24	59 ± 22	1.6×10^{-12}	2.3×10^{-8}	0.0025
eejs15	11	0.78	0.22	0.29	256	69	102.3	$(-1.08 \pm 0.35) \times 10^2$	-87 ± 31	-65 ± 27	2.9×10^{-13}	7.9×10^{-7}	0.0069
eejs15	12	0.73	0.26	0.36	298	87	105.6	26 ± 18	21 ± 16	16 ± 14	1.3×10^{-10}	5.5×10^{-6}	0.034
eejs15	13*	1.30	0.33	0.25	525	126	199.8	$93 \pm 1.5 \times 10^2$	$74 \pm 1.3 \times 10^2$	$55 \pm 1.1 \times 10^2$	7.1×10^{-8}	5.4×10^{-6}	0.0043
eejs15	14	0.50	0.18	0.36	170	53	33.1	-26 ± 75	-21 ± 65	-16 ± 56	0.14	0.13	0.24
eejs15	15	0.50	0.13	0.26	234	61	32.0	-73 ± 56	-58 ± 48	-44 ± 43	2.7×10^{-8}	1.8×10^{-7}	0.062
eejs15	16	0.67	0.33	0.49	213	120	145.0	$(1.37 \pm 0.51) \times 10^2$	$(1.10 \pm 0.45) \times 10^2$	83 ± 42	2.4×10^{-5}	0.013	0.43
eejs15	17*	1.15	0.41	0.36	177	69	168.3	$(7.9 \pm 1.3) \times 10^2$	$(6.3 \pm 1.1) \times 10^2$	$(4.75 \pm 0.93) \times 10^2$	3.5×10^{-19}	1.2×10^{-10}	0.0028
ejs15	18	0.81	0.36	0.44	63	55	76.3	-81 ± 48	-65 ± 42	-49 ± 37	1.5×10^{-5}	0.0039	0.29
jsres	19	1.04	0.32	0.31	166	89	79.5	55 ± 22	44 ± 19	33 ± 17	3.0×10^{-10}	1.5×10^{-6}	0.0078
jsres	20	1.27	0.63	0.50	134	89	176.8	-84 ± 29	-67 ± 25	-50 ± 23	5.6×10^{-7}	0.011	0.50
Measured								Δ^{17}O (p.p.m.)					
Earth		1						0 ± 3					
Moon		0.012						12 ± 3					
Mars		0.07						321 ± 13					
4 Vesta		4.33×10^{-5}						-250 ± 80					

M_P, N_P and M_I, N_I are the mass and number of particles in the planet and the impactor, respectively; t_{coll} is the collision time in the simulations; C_{cal} is the calibration pre-factor (see main text). The Δ^{17}O composition difference and the Kolmogorov–Smirnov probability (of the distribution of the feeding zones of the planet and impactor being sampled from the same parent distribution) are shown for both the case of no contribution of planetary material to the newly formed Moon, and the cases of 20% and 40% contribution of material from the planet.
* Three-planet systems; calibration was done on the second and third planets.

discussed here, as shown in the Methods. The comparison is then done as follows. First we compare the feeding zones of the planet and impactor, as shown in the examples in Fig. 1 (the cumulative plots for these and all other cases can be found in the Methods). We calculate the probability P that the feeding zones of the impactor and the planet are drawn from the same distribution, using a two-group Kolmogorov–Smirnov test (probabilities shown in the plots and in Table 1). In 3 out of 20 cases the feeding zones contributing to the Moon and those contributing to the planet are consistent with being drawn from the same parent distribution. In other words, the Moon's feeding zones, if derived solely from the impactor, are consistent with the Earth's in 15% of the impacts. The consistency further improves if we assume that a fraction of the proto-Earth was mixed into the Moon (as suggested by detailed collision simulations showing a 10%–40% contribution from the proto-Earth[14]). For the typical 20% mix of proto-Earth material with the impactor material forming the Moon (as found in simulations), 35% of cases are consistent with their feeding zones being drawn from the same parent distribution, and the success rate increases further for a

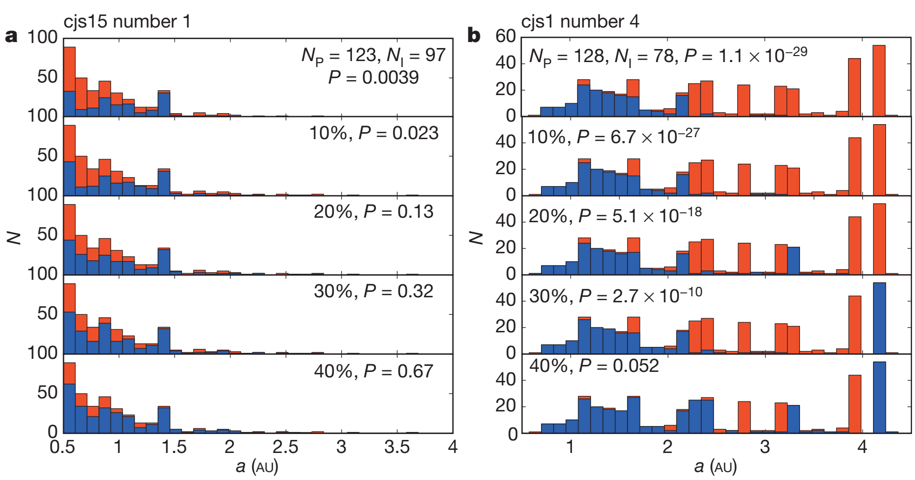


Figure 1 | The distribution of planetesimals composing the planet and the impactor. **a**, A case where the origins of the planetesimals composing the planet (red) and the impactor (blue) are consistent with being sampled from the same parent distribution for the expected typical 20% contribution of planetary material in moon-forming impacts (Kolmogorov–Smirnov test probability >0.05). **b**, A case where the planet and impactor compositions are inconsistent ($P < 0.05$), but become consistent once a large (40%) contribution of material from the planet is considered. The lower plots in each panel show the results when different contributions from the planet are assumed (four cases are shown 10%; 20%; 30% and 40%). The cumulative distribution for these cases as well as all other planet–impactor pairs in Table 1 can be found in the Methods.

higher mass contribution from the proto-Earth (see Table 1 and Extended Data Figs 1 and 2). While this shows that the proto-Earth and the Moon-forming impactor may have had similar feeding zones, it does not yet quantitatively guarantee that the composition is as similar as that of the Earth–Moon system.

We therefore further explore the compositional similarities with the Earth–Moon system, and calculate the oxygen isotope ratios of our simulated planets. We assume that a linear gradient existed in the ^{17}O isotopic composition in the initial protoplanetary disk of the Solar System. Following ref. 21 we calibrate the initial ^{17}O isotopic composition in each of our simulations using the measured compositions of Earth and Mars (see ref. 21 and the Methods, where we also discuss the sensitivity of the results to the calibration used, as well as the criteria for which planet–impactor pairs are considered in the analysis. We find qualitatively similar results when using different criteria and calibrations, as we discuss in detail in the Methods). Given this calibration we assign each planetesimal and planetary embryo a specific initial ^{17}O isotopic abundance based on its initial orbit, and then average the contribution of all accreted planetesimals, while weighting each accreted planetesimal/embryo according to its mass, to obtain the ^{17}O of the planets and impactors and derive the offset between them.

The ^{17}O isotope is chosen for comparison since it provides the most stringent constraint on the Earth–Moon similarity, and its abundances were measured across a variety of Solar System bodies (enabling the best opportunity for calibration). The measured difference between the ^{17}O abundances of Mars and Earth (used for calibration) is $\Delta^{17}O_{Mars} = +321 \pm 13$ parts per million (p.p.m.)[12] (a similarly large difference was found for the composition of 4 Vesta asteroid derived from HED (howardite–eucrite–diogenite) meteorites; -250 ± 80 p.p.m.; ref. 13). The difference between the Earth and Moon is just $\Delta^{17}O = 12 \pm 3$ p.p.m. (ref. 11). In Table 1 we show the Δ^{17}O differences between the impactors and the planets in our simulations. We adopt the same Earth–Mars composition difference calibration as used by ref. 21. The results are linearly dependent on the adopted calibration; see Table 1. The calibration factor is defined as $C_{cal} = (\Delta^{17}O_4 - \Delta^{17}O_3)/\Delta^{17}O_{Mars}$, where $\Delta^{17}O_4$ and $\Delta^{17}O_3$ refer to the compositions of the fourth and third planet in the simulations (unless only three planets formed, in which case the third and second planet were taken for calibration purposes).

In 20% of the cases the impactors and planets have absolute offsets comparable or smaller than the measured absolute offset in the Earth–Moon system, that is, smaller than the 1σ limit estimated using the lunar samples (<15 p.p.m.). Taking into account the 1σ uncertainty calculated for the Δ^{17}O in the simulated systems, the fraction of consistent pairs can be as large as 40% of the whole sample. This fraction becomes larger when partial mixing of Earth material is allowed, as observed in simulation data (it increases to 50% for a 20% (and to 55% for a 40%) contribution from the planet; see Table 1 and Fig. 2). Even planet–impactor pairs with statistically different feeding zones have Δ^{17}O offsets significantly (see Table 1) smaller than those found for Mars and Vesta, in most cases. More generally, planet–impactor pairs are robustly more similar in composition than are pairs of surviving planets in the same system (see Fig. 2 and Extended Data Figs 3, 4 and 5, as well as the Supplementary Table, for the Δ^{17}O difference distribution for planets and impactors). No less importantly, the differences between the planets are of a similar order to those found between the Earth, Mars and Vesta, that is, consistent with the observations of the Solar System. Interestingly, a small fraction of the planets do have very similar composition (small Δ^{17}O difference), suggesting the possibility of the existence of Solar System bodies with similar compositions to the Earth besides the Moon. As shown in Extended Data Table 2 and in the Methods, this result still holds when considering lower thresholds for the minimal number of particles composing the planet or impactor (between 1 to 40). In particular, the mean fraction of compatible planet–impactor pairs falls between 10% and 20% for all cases. The fractions become even higher (20%–40%) when accounting for the 1σ uncertainties.

The Earth–Moon composition similarity poses a major challenge to the standard model of the giant-impact scenario because it conflicts with the predominant derivation of the Moon composition from the mantle of the impacting planet[14]. A wide range of alternative impact scenarios were therefore investigated[14,16,17,21–24]. However, all such models suffer from potentially considerable difficulties and/or require fine-tuned conditions (see ref. 14 for a review). Our analysis of Solar-System-like planet formation scenarios potentially offers a solution to the major composition-similarity obstacle to the standard giant-impact scenario. We find that a significant fraction of all planetary impactors could have had compositions similar to the planets they struck, in contrast with the distinct compositions of different planets existing in the same planetary system.

Note that the solution, suggested by our results, of the impactor and planet having a similar composition is also applicable to the origin of the Δ^{17}O similarity between the Earth and the Moon, and may similarly apply for the isotopic similarities of silicon and tungsten[25]. However, it

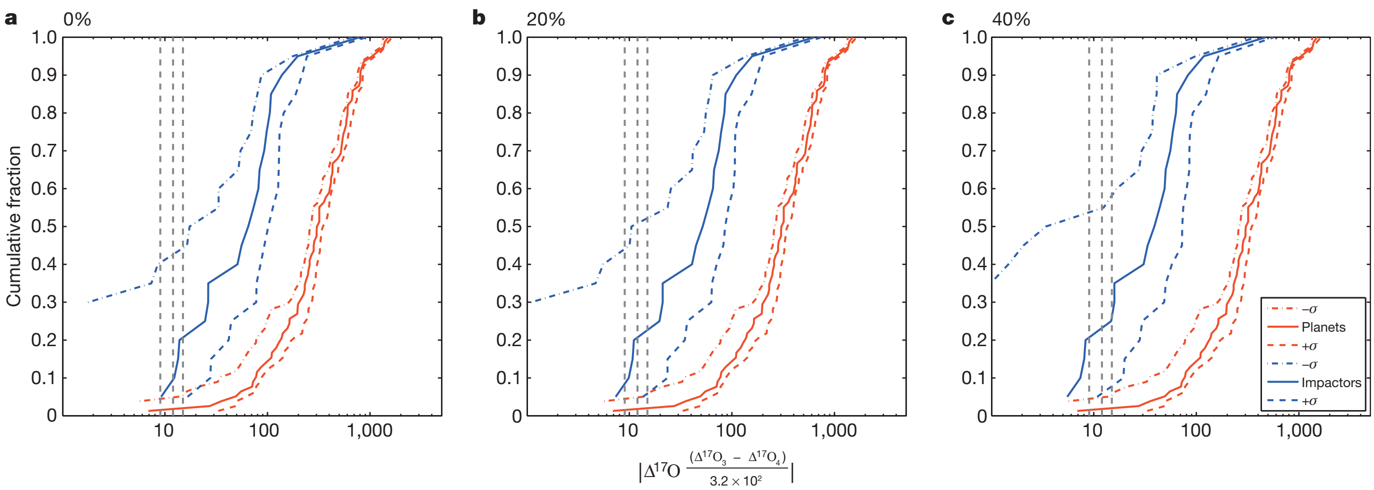


Figure 2 | The cumulative distribution of the absolute Δ^{17}O differences between planets and their last giant impactors (blue), compared with the differences between planets in the same system (red). Panels **a**, **b** and **c** correspond to the cases of zero, 20% and 40% contribution of material from the planet to a Moon formed from these impacts, respectively. The vertical lines depict the Δ^{17}O difference of the Earth–Moon system (dashed lines for $\pm\sigma$ around the mean value; central dashed line). The differences between the planet–impactor pairs are systematically smaller than those found between different planets (the same parent distribution for the two groups can be excluded with high confidence; Kolmogorov–Smirnov probability 6.7×10^{-8}, 1.1×10^{-8} and 1.3×10^{-9} for the zero, 20% and 40% cases, respectively).

is debatable whether even a similar impactor–planet composition could resolve the compositional similarity of silicon, given that the Earth's silicate mantle may reflect the consequences of silicon sequestration by a core formed at high temperatures on a large planetary body[25,26], that is, larger than the typical impactors considered.

We conclude that our findings can potentially resolve the apparent contrast between the observed similarity of the Earth and the Moon composition and its difference from that of other Solar System bodies. This primordial composition similarity solution may therefore remove the main obstacle to the standard giant-impact origin of the Moon, as well as ease some of the difficulties for the alternative giant-impact scenarios suggested in recent years[14].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 November 2014; accepted 10 February 2015.

1. Canup, R. M. & Asphaug, E. Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* **412,** 708–712 (2001).
2. Agnor, C. B., Canup, R. M. & Levison, H. F. On the character and consequences of large impacts in the late stage of terrestrial planet formation. *Icarus* **142,** 219–237 (1999).
3. Jacobson, S. A. & Morbidelli, A. Lunar and terrestrial planet formation in the Grand Tack scenario. *Phil. Trans. R. Soc. Lond. A* **372,** 20130174 (2014).
4. Canup, R. M. Simulations of a late lunar-forming impact. *Icarus* **168,** 433–456 (2004).
5. Canup, R. M. Lunar-forming collisions with pre-impact rotation. *Icarus* **196,** 518–538 (2008).
6. Ringwood, A. E. Terrestrial origin of the moon. *Nature* **322,** 323–328 (1986).
7. Lugmair, G. W. & Shukolyukov, A. Early solar system timescales according to ^{53}Mn-^{53}Cr systematics. *Geochim. Cosmochim. Acta* **62,** 2863–2886 (1998).
8. Wiechert, U. *et al.* Oxygen isotopes and the Moon-forming giant impact. *Science* **294,** 345–348 (2001).
9. Touboul, M., Kleine, T., Bourdon, B., Palme, H. & Wieler, R. Late formation and prolonged differentiation of the Moon inferred from W isotopes in lunar metals. *Nature* **450,** 1206–1209 (2007).
10. Zhang, J., Dauphas, N., Davis, A. M., Leya, I. & Fedkin, A. The proto-Earth as a significant source of lunar material. *Nature Geosci.* **5,** 251–255 (2012).
11. Herwartz, D., Pack, A., Friedrichs, B. & Bischoff, A. Identification of the giant impactor Theia in lunar rocks. *Science* **344,** 1146–1150 (2014).
12. Franchi, I. A., Wright, I. P., Sexton, A. S. & Pillinger, C. T. The oxygen-isotopic composition of Earth and Mars. *Meteorit. Planet. Sci.* **34,** 657–661 (1999).
13. Clayton, R. N. & Mayeda, T. K. Oxygen isotope studies of achondrites. *Geochim. Cosmochim. Acta* **60,** 1999–2017 (1996).
14. Asphaug, E. Impact origin of the Moon? *Annu. Rev. Earth Planet. Sci.* **42,** 551–578 (2014).
15. Raymond, S. N., O'Brien, D. P., Morbidelli, A. & Kaib, N. A. Building the terrestrial planets: constrained accretion in the inner Solar System. *Icarus* **203,** 644–662 (2009).
16. Ćuk, M. & Stewart, S. T. Making the Moon from a fast-spinning Earth: a giant impact followed by resonant despinning. *Science* **338,** 1047–1052 (2012).
17. Canup, R. M. Forming a Moon with an Earth-like composition via a giant impact. *Science* **338,** 1052–1055 (2012).
18. Chambers, J. E. A hybrid symplectic integrator that permits close encounters between massive bodies. *Mon. Not. R. Astron. Soc.* **304,** 793–799 (1999).
19. Raymond, S. N., Kokubo, E., Morbidelli, A., Morishima, R. & Walsh, K. J. Terrestrial planet formation at home and abroad. In *Protostars and Planets VI* (eds Beuther, H., Klessen, R., Dullemond, C. & Henning, Th.), 585–618 (University of Arizona Press, 2014).
20. Raymond, S. N., Quinn, T. & Lunine, J. I. High-resolution simulations of the final assembly of Earth-like planets. I. Terrestrial accretion and dynamics. *Icarus* **183,** 265–282 (2006).
21. Pahlevan, K. & Stevenson, D. J. Equilibration in the aftermath of the lunar-forming giant impact. *Earth Planet. Sci. Lett.* **262,** 438–449 (2007).
22. Belbruno, E. & Gott, J. R. III. Where did the Moon come from? *Astron. J.* **129,** 1724–1745 (2005).
23. Salmon, J. & Canup, R. M. Lunar accretion from a Roche-interior fluid disk. *Astrophys. J.* **760,** 83 (2012).
24. Reufer, A., Meier, M. M. M., Benz, W. & Wieler, R. A hit-and-run giant impact scenario. *Icarus* **221,** 296–299 (2012).
25. Dauphas, N., Burkhardt, C., Warren, P. & Teng, F.-Z. Geochemical arguments for an Earth-like Moon-forming impactor. *Phil. Trans. R. Soc. Lond. A* **372,** 20130244 (2014).
26. Elliott, T. & Stewart, S. T. Planetary science: shadows cast on Moon's origin. *Nature* **504,** 90–91 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements H.B.P. acknowledges support from BSF grant number 2012384, the Minerva Center for Life under Extreme Planetary Conditions, the ISF I-CORE grant number 1829/12 and the Marie Curie IRG 333644 'GRAND' grant. S.N.R. acknowledges funding from the Agence Nationale pour la Recherche via grant ANR-13-BS05-0003-002 (project MOJO). We thank O. Aharonson for remarks on an early version of this manuscript. We thank N. Kaib and N. Cowan for helpful discussions on their related work.

Author Contributions A.M.-B. analysed the simulation data and produced the main results; H.B.P. initiated and supervised the project and took part in the data analysis. S.N.R. provided the simulation data used for the analysis. The paper was written by A.M.-B. and H.B.P. with contributions from S.N.R.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M.B. (amastrobuono@physics.technion.ac.il) and H.B.P. (hperets@physics.technion.ac.il).

METHODS

In the following we supply additional data on the compositions of planet–impactor pairs and the compositions of different planets at the same systems. We also provide more detailed information on the methods used, as well as discuss the sensitivity of our results to the various criteria and calibrations which we applied. The full simulation data used in this work can be provided by the authors upon request.

Initial configuration of Jupiter and Saturn. The simulations analysed here are described in detail in ref. 15. The various simulations explore a range of different initial conditions for the gaseous planets. In particular, in the cjs and cjsecc simulations Jupiter and Saturn are placed on orbits with semimajor axes of 5.45 AU and 8.18 AU and mutual inclination of 0.5°. In cjs the orbits are circular while in cjsecc they are eccentric with $e_{\mathrm{Jupiter}} = 0.02$ and $e_{\mathrm{Saturn}} = 0.03$. In eejs, Jupiter and Saturn are placed in their current positions (5.25 AU and 9.54 AU), with mutual inclination of 1.5° and larger eccentricities than the observed ones ($e_{\mathrm{Jupiter}} = e_{\mathrm{Saturn}} = 0.1$ or $e_{\mathrm{Jupiter}} = 0.07$ and $e_{\mathrm{Saturn}} = 0.08$). In ejs the orbits of Jupiter and Saturn have similar parameters to those observed ($a_{\mathrm{Jupiter}} = 5.25$AU and $e_{\mathrm{Jupiter}} = 0.05$, $a_{\mathrm{Saturn}} = 9.54$ and $e_{\mathrm{Saturn}} = 0.06$) with mutual inclination of 1.5°. Finally, in jsres, Jupiter and Saturn are placed at $a_{\mathrm{Jupiter}} = 5.43$ AU and $a_{\mathrm{Saturn}} = 7.30$ AU with $e_{\mathrm{Jupiter}} = 0.005$ and $e_{\mathrm{Saturn}} = 0.01$ and with mutual inclination of 0.2°.

The composition difference between planets in the same system. The Supplementary Table shows the Δ^{17}O differences between the different planets (in each system) and the impacted planets analysed in the main text. The full cumulative distribution for these data can be seen in Fig. 2. Note that in the cases where two impacted planets were analysed in the same system, the differences are shown with respect to both the first and second planets.

The cumulative composition distribution of planet–impactor pairs. We used the following procedure to calculate the spatial distribution of the feeding zones (used for Fig. 1, Extended Data Figs 1 and 2 and the Kolmogorov–Smirnov probabilities in Table 1). We extracted the record of planetesimals that constitute the planet and the impactor before the last Moon-forming impact, as well as the planet composition after the collision. To account for the different contributions of particles of different masses we replicated n_i times each particle, where n_i is the ratio between the mass of the ith particle and the minimum mass of the planetesimals. The planetesimal record was then used to produce the distribution of the feeding zones used in our analysis. In cases where contribution of planetary material to the Moon composition was considered, we randomly chose particles from the planet and added them to the planetesimals composing the impactor, where appropriate numbers of particles were taken so as to produce the relevant fractional contribution (for the cases of 10%, 20%, 30% or 40% contribution). We then repeated the same analysis as done for the impactor, with these new mixed impactors.

Δ^{17}O calibration. To calculate the Δ^{17}O ($\equiv \delta^{17}\mathrm{O} - 0.52\delta^{16}\mathrm{O}$) for the planet and impactor pairs we followed the procedure described by ref. 21. To assign specific values of Δ^{17}O to each particle in the simulation we calibrated our simulations with the Solar System observations. We assume a linear gradient of Δ^{17}O with heliocentric distance r

$$\Delta^{17}\mathrm{O}(r) = c_1 r + c_2 \tag{1}$$

where the two free parameters in equation (1) were calibrated by assuming that the third planet formed in the system has the composition of the Earth ($\Delta^{17}\mathrm{O} = 0‰$) and the fourth one has the composition of Mars ($\Delta^{17}\mathrm{O} = +0.32‰$). In cases where only three planets formed (marked with an asterisk in Table 1), we assigned the composition of the Earth and of Mars to the second and third planets in the simulation, respectively. We then mass-averaged over all the $\Delta^{17}\mathrm{O}(r)$ of the planetesimals that accreted to form the Earth. We used the heliocentric distance r as the initial position of each body. We did the same for the planetesimals composing the simulated Mars. In this way we have the system of equations

$$\begin{cases} \dfrac{\sum^{N} m_{i,\oplus}(c_1 r_{i,\oplus} + c_2)}{M_\oplus} = 0 \\ \dfrac{\sum^{N'} m_{i,\mathrm{Mars}}(c_1 r_{i,\mathrm{Mars}} + c_2)}{M_{\mathrm{Mars}}} = 0.32 \end{cases}$$

where $r_{i,\oplus}$ and $m_{i,\oplus}$ ($r_{i,\mathrm{Mars}}$ and $m_{i,\mathrm{Mars}}$) are the initial position and mass of the ith planetesimal composing the Earth (Mars) and $M_\oplus$ (M_{Mars}) is the final total mass of the Earth (Mars). In this way it has been possible to evaluate the Δ^{17}O value for each planetesimal in the system and thus for all the planets in each system. Given this calibration we evaluate the Δ^{17}O of the planet ($\Delta^{17}\mathrm{O_P}$) and of the last impactor ($\Delta^{17}\mathrm{O_I}$), as the average of the Δ^{17}O of all their respective components, as well as calculated the 1σ s.e.m. for each of these values (σ_{P} and σ_{I}). To check whether or not the planet–Moon system is consistent with the Earth–Moon system we evaluated the difference $\Delta^{17}\mathrm{O_I} - \Delta^{17}\mathrm{O_P}$ and the relative error $\sigma = \sqrt{\sigma_{\mathrm{P}}^2 + \sigma_{\mathrm{I}}^2}$.

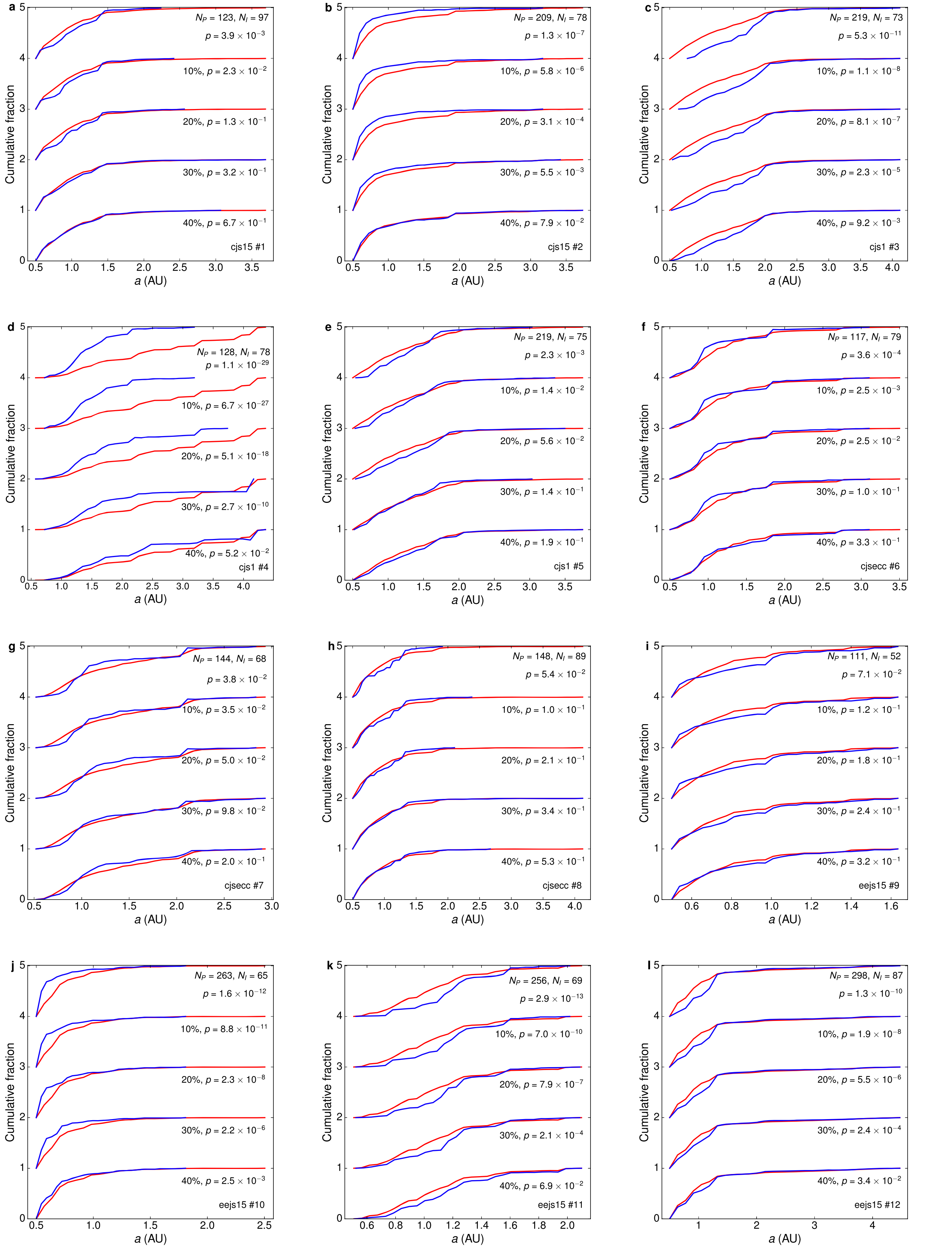
To calculate the Δ^{17}O when a fractional contribution from the planet is included, we added the average Δ^{17}O of the planet and the impactor, each weighted according to the appropriate fractional contribution considered.

To study the sensitivity of our results to the Earth–Mars composition difference calibration used, we also considered lower and higher calibrations, between 0.5 and 1.5 times the Earth–Mars Δ^{17}O difference. We re-analysed the fraction of compatible planet–impactor pairs (that is, producing planet–moon pairs with a composition difference equal or smaller than the Earth–Moon composition difference) for these different calibration factors. The results are summarized in Extended Data Table 1. We find that although the fraction of consistent pairs decreases with the use of a larger difference calibration, as expected, difference calibrations as much as 1.5 times larger than the Earth–Mars difference still give rise to a mean 5% of planet–impactor pairs with similar composition (and 40% within 1σ uncertainty in the simulation compositions), rising to a mean 10% to 20% for the cases of 20% and 40% mixing of the planetary material, respectively. In other words, the results are generally robust to that level and do not dependent on a fine-tuned calibration.

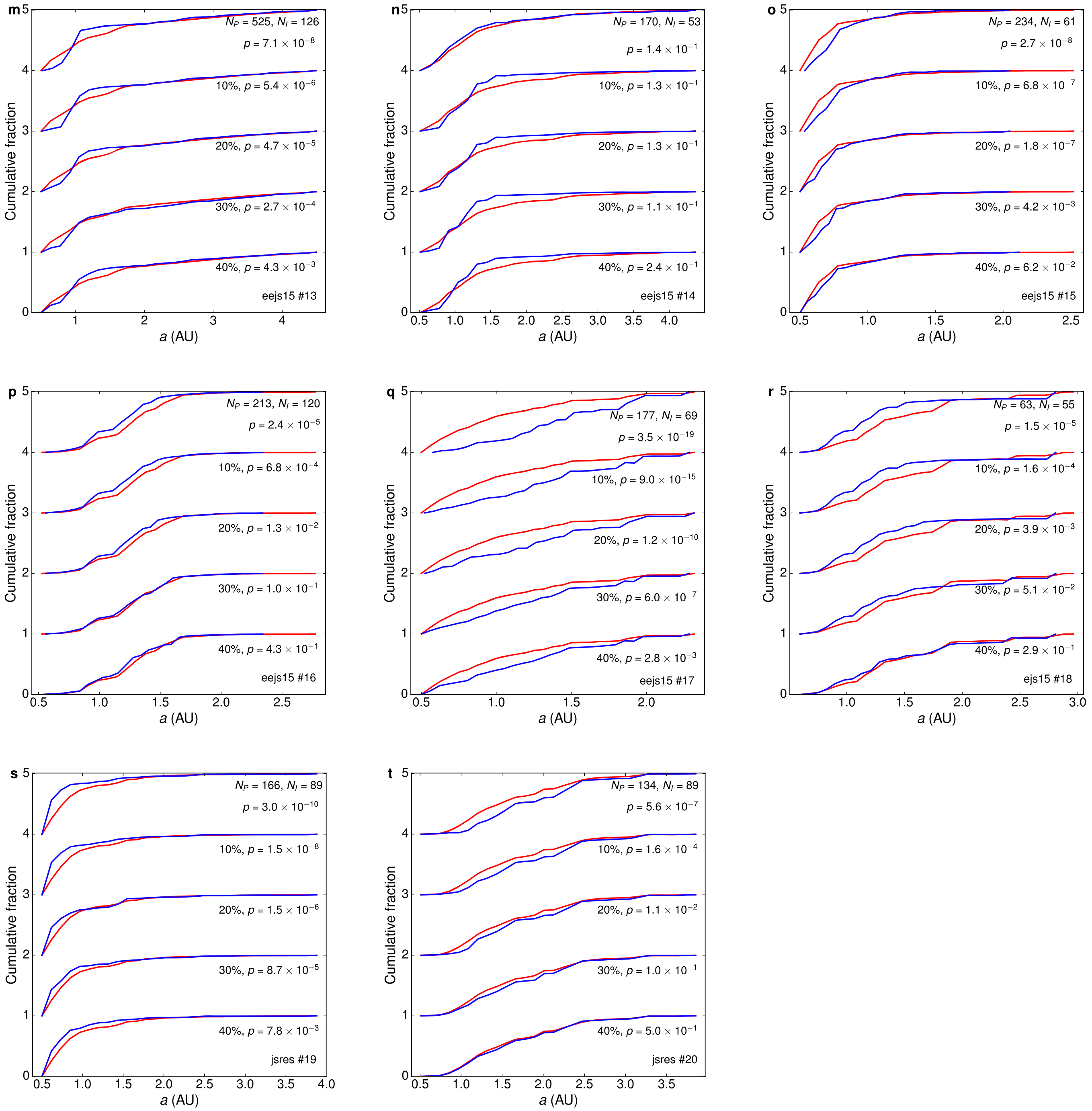
Dependence on the criteria for the chosen planet–impactor pairs. To verify the robustness of our results to different criteria for the choice of planet–impactor pairs used in our analysis we studied two different sets of criteria:

(1) Use all planet–last-impactor pairs, considering smaller thresholds (that is, smaller than the 50-particles threshold considered in the main text) for the number of composing particles (and corresponding masses), but requiring an impactor mass of at least $0.5M_{\mathrm{Mars}}$ to ensure a Moon-forming impact. Taking a threshold of 1, 10, 20, 30 and 40 as the minimal number of particles, we find that the general conclusion is unchanged; impactors have compositions more similar to the planets they impact than to other planets in the system. The mean fraction of planet–impactor pairs with similarity comparable to that of the Earth–Moon system is between 10% and 20% in all cases (and up to 40% when considering the 1σ uncertainties), as shown in Extended Data Table 2. Extended Data Figs 3, 4 and 5 show the cumulative distributions of the compositions of planets and last impactors for all the systems, regardless of the number of particles that contributed to their formation, and for a minimum of 10, 20, 40 and 50 particles composing the planet and last impactor. Extended Data Figs 3, 4 and 5 are shown for 0%, 20% and 40% mixing between the material of the planet and impactor.

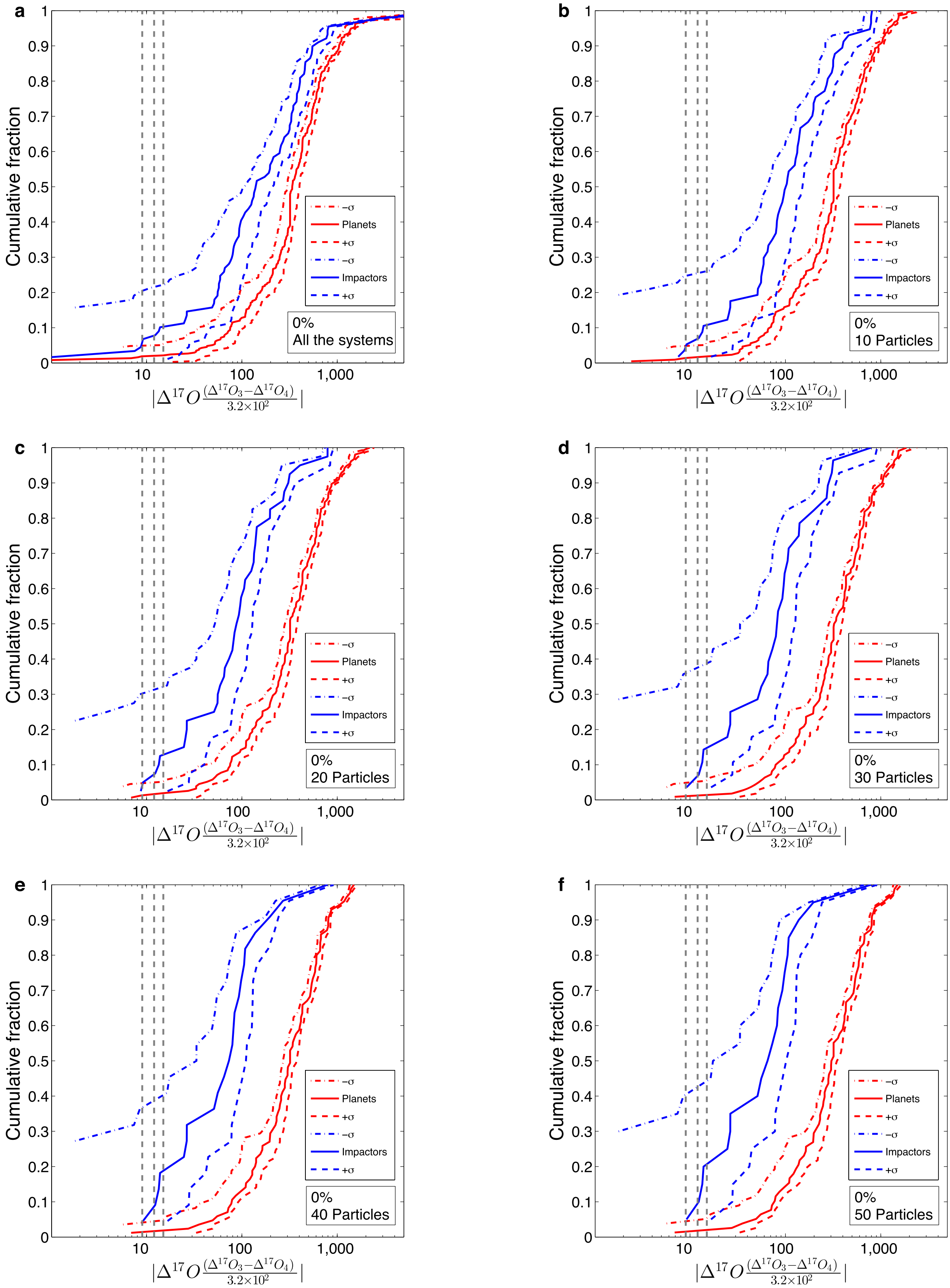
(2) Consider only last impactors on the third planet. Once we require the impactor to have at least $0.5M_{\mathrm{Mars}}$ and be composed of a significant (>20) number of particles, the statistics become too small. When we do not consider a minimal threshold for the number of composing particles, we find that 2 out of 18 (11%; up to about 30% with the 1σ uncertainty in the simulation compositions) planet–impactor pairs with a composition difference equal or smaller than the Earth–Moon system.



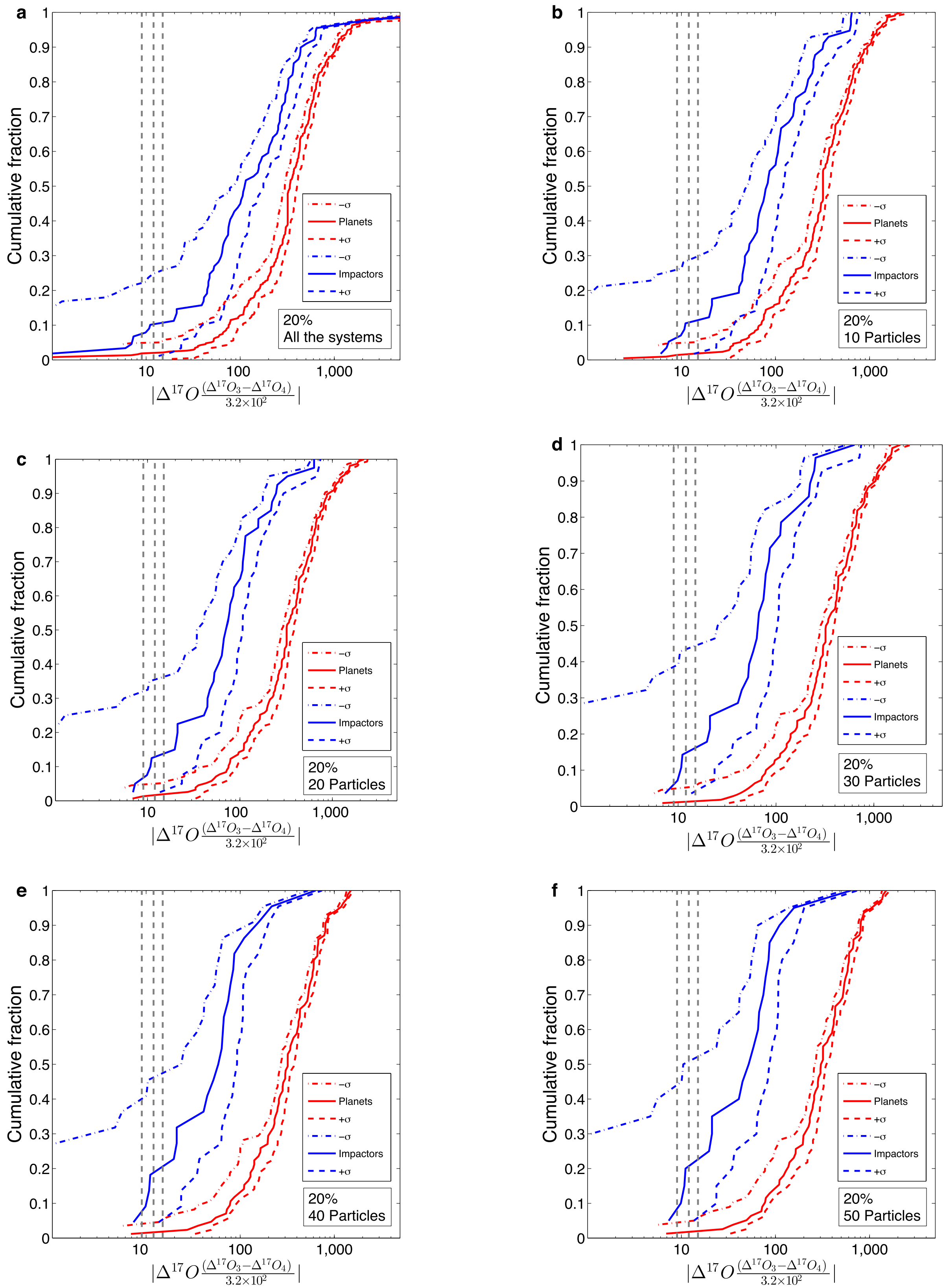
Extended Data Figure 1 | The cumulative distribution of the planetesimals composing the planet (red) and the impactor (blue). All planet–impactor pairs in Table 1 are shown, cases 1–12 (panels **a–l**), including the cumulative distributions corresponding to the histograms in Fig. 1.



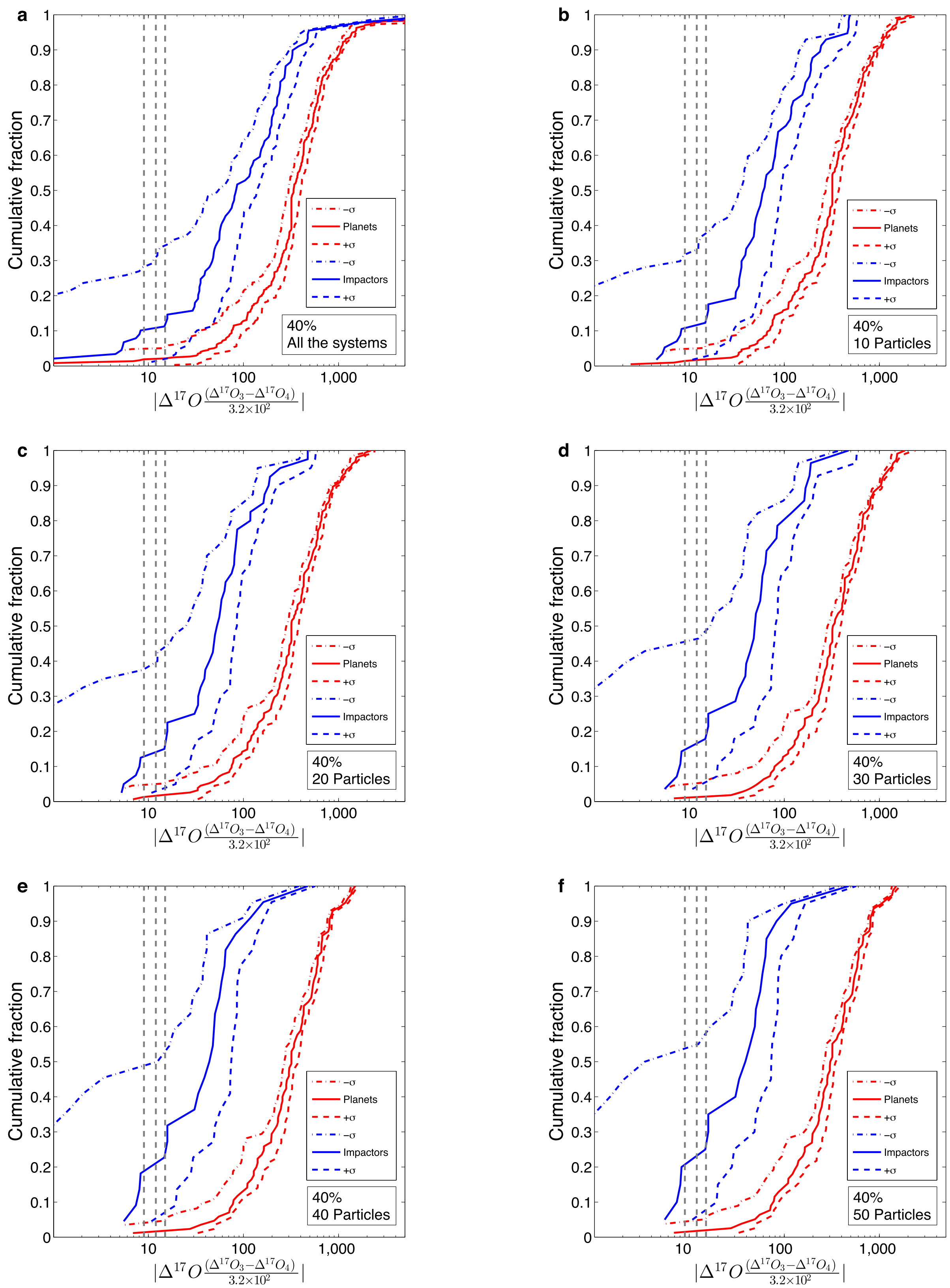
Extended Data Figure 2 | The cumulative distribution of the planetesimals composing the planet (red) and the impactor (blue). All planet–impactor pairs in Table 1 are shown, cases 13–20 (panels **m**–**t**).



Extended Data Figure 3 | The cumulative distributions of the compositions of planets and last impactors assuming 0% mixing between Earth and Moon material. The cumulative distribution of the absolute $\Delta^{17}O$ differences between planets and their last giant impactors (blue), compared with the differences between planets in the same system (red), assuming 0% mixing between Earth and Moon material. From the top left panel (**a**) to the bottom right panel (**f**) we consider all the systems, regardless of the number of particles that contributed to their formation, and planets and last impactors composed of a minimum of 10, 20, 40 and 50 particles. Only last impactors with mass $>0.5M_{\mathrm{Mars}}$ have been taken into account.



Extended Data Figure 4 | The cumulative distributions of the compositions of planets and last impactors assuming 20% mixing between Earth and Moon material. The cumulative distribution of the absolute Δ^{17}O differences between planets and their last giant impactors (blue), compared with the differences between planets in the same system (red), assuming 20% mixing between Earth and Moon material. From the top left panel (**a**) to the bottom right panel (**f**) we consider all the systems, regardless of the number of particles that contributed to their formation, and planets and last impactors composed of a minimum of 10, 20, 40 and 50 particles. Only last impactors with mass $>0.5M_{\mathrm{Mars}}$ have been taken into account.



Extended Data Figure 5 | The cumulative distributions of the compositions of planets and last impactors assuming 40% mixing between Earth and Moon material. The cumulative distribution of the absolute Δ^{17}O differences between planets and their last giant impactors (blue), compared with the differences between planets in the same system (red), assuming 40% mixing between Earth and Moon material. From the top left panel (**a**) to the bottom right panel (**f**) we consider all the systems, regardless of the number of particles that contributed to their formation, and planets and last impactors composed of a minimum of 10, 20, 40 and 50 particles. Only last impactors with mass $>0.5M_{\mathrm{Mars}}$ have been taken into account.

Extended Data Table 1 | The mean fraction of last impactors with a composition compatible with the planet they impact

Factor	0%		20%		40%	
	Mean	1σ	Mean	1σ	Mean	1σ
0.5	35%	50%	35%	60%	35%	70%
0.75	20%	50%	25%	50%	35%	60%
1	20%	40%	20%	50%	25%	55%
1.25	5%	40%	20%	50%	20%	50%
1.5	5%	40%	10%	40%	20%	50%

The fraction is given for different calibration factors and mixing percentages. The pairs which are consistent within 1σ of the simulation uncertainties are also given.

Extended Data Table 2 | The mean fraction of planet–impactor consistent pairs

N_{min}	0%		20%		40%		N_{cases}
	Mean	1σ	Mean	1σ	Mean	1σ	
0	10.1%	21.3%	10.1%	25.8%	11.2%	33.7%	89
10	10.5%	24.6%	10.5%	29.8%	12.3%	36.9%	57
20	12.5%	30%	12.5%	35%	15%	42.5%	40
30	14.3%	35.7%	14.3%	42.9%	17.9%	46.4%	28
40	18.2%	36.4%	18.2%	45.5%	22.7%	50%	22
50	20%	40%	20%	50%	25%	55%	20

The fraction is shown for different mixing percentages (0%, 20%, 40%) and minimum numbers (N_{min}) of the particles composing the impactor and planet (1, 10, 20, 30, 40, 50). The pairs which are consistent within 1σ of the simulation uncertainties are also given. Only the number of cases N_{cases} in which the last impactor has a mass $>0.5M_{Mars}$ have been taken into account.

doi:10.1038/nature14134

Levantine cranium from Manot Cave (Israel) foreshadows the first European modern humans

Israel Hershkovitz[1,2]*, Ofer Marder[3]*, Avner Ayalon[4], Miryam Bar-Matthews[4], Gal Yasur[4], Elisabetta Boaretto[5], Valentina Caracuta[5], Bridget Alex[5,6], Amos Frumkin[7], Mae Goder-Goldberger[8], Philipp Gunz[9], Ralph L. Holloway[10], Bruce Latimer[11,12], Ron Lavi[13], Alan Matthews[14], Viviane Slon[1]†, Daniella Bar-Yosef Mayer[2], Francesco Berna[15], Guy Bar-Oz[16], Reuven Yeshurun[16], Hila May[2,17], Mark G. Hans[12], Gerhard W. Weber[18,19] & Omry Barzilai[20]

A key event in human evolution is the expansion of modern humans of African origin across Eurasia between 60 and 40 thousand years (kyr) before present (BP), replacing all other forms of hominins[1]. Owing to the scarcity of human fossils from this period, these ancestors of all present-day non-African modern populations remain largely enigmatic. Here we describe a partial calvaria, recently discovered at Manot Cave (Western Galilee, Israel) and dated to 54.7 ± 5.5 kyr BP (arithmetic mean ± 2 standard deviations) by uranium–thorium dating, that sheds light on this crucial event. The overall shape and discrete morphological features of the Manot 1 calvaria demonstrate that this partial skull is unequivocally modern. It is similar in shape to recent African skulls as well as to European skulls from the Upper Palaeolithic period, but different from most other early anatomically modern humans in the Levant. This suggests that the Manot people could be closely related to the first modern humans who later successfully colonized Europe. Thus, the anatomical features used to support the 'assimilation model' in Europe might not have been inherited from European Neanderthals, but rather from earlier Levantine populations. Moreover, at present, Manot 1 is the only modern human specimen to provide evidence that during the Middle to Upper Palaeolithic interface, both modern humans and Neanderthals contemporaneously inhabited the southern Levant, close in time to the likely interbreeding event with Neanderthals[2,3].

Manot is an active karstic cave located 40 km northeast of the Mount Carmel cave sites (Fig. 1). Archaeological material retrieved in five excavation seasons (2010–2014) currently indicates that the cave was intensively occupied during the early Upper Palaeolithic period[4], and, to a lesser extent, during the Initial Upper Palaeolithic and late Middle Palaeolithic periods (Supplementary Information A and Extended Data Figs 1 and 2). The original cave entrance was blocked following the collapse of the roof, probably between 30 kyr and 15 kyr ago[4]. The archaeological evidence from the cave indicates two major cultural events. The first is associated with the earlier dispersal of anatomically modern humans (AMHs) (Middle Palaeolithic), best represented in the Near East by the Qafzeh and Skhul fossils (~120–90 kyr ago)[5]. The second corresponds to the colonization of the Eastern Mediterranean region by 'modern humans' ~45 kyr ago (Upper Palaeolithic), presumably replacing the Neanderthals (for example, fossils from Amud, Kebara and Dederiyeh) in the region (~65–50 kyr ago)[6].

Figure 1 | Geographical location of Manot Cave, Israel. Middle Palaeolithic and Upper Palaeolithic sites with human remains are marked. Manot 1 is penecontemporaneous with the nearby Neanderthals of Amud and Kebara Caves (insert map) and older than all European Upper Palaeolithic specimens (large map). The nearby sites of Skhul, Kebara and Tabun Caves are situated on the western slope of Mount Carmel. The modern city of Haifa is shown (black triangle).

[1]The Dan David Laboratory for the Search and Study of Modern Humans, Sackler Faculty of Medicine, Tel Aviv University, PO Box 39040, Tel Aviv 6997801, Israel. [2]The Steinhardt Museum of Natural History and National Research Center, Tel Aviv University, PO Box 39040, Tel Aviv 6997801, Israel. [3]Archaeology Division, Ben-Gurion University of the Negev, PO Box 653, Beer-Sheva 8410501, Israel. [4]Geological Survey of Israel, 30 Malkhe Israel Street, Jerusalem 95501, Israel. [5]Max Planck Society-Weizmann Institute Center for Integrative Archaeology and Anthropology, D-REAMS Radiocarbon Laboratory, Weizmann Institute of Science, Rehovot 76100, Israel. [6]Department of Anthropology and Human Evolutionary Biology, Harvard University, 11 Divinity Avenue, Cambridge, Massachusetts 02138, USA. [7]Department of Geography, The Hebrew University of Jerusalem, Jerusalem 91905, Israel. [8]Institute of Archaeology, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel. [9]Department of Human Evolution, Max-Planck-Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103, Leipzig, Germany. [10]Department of Anthropology, Columbia University, New York 10027, USA. [11]Department of Anatomy, Case Western Reserve University, Cleveland, Ohio 44106, USA. [12]Department of Orthodontics, Case Western Reserve University School of Dental Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA. [13]8 Dan Street, Modi'in 7173161, Israel. [14]Institute of Earth Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem 91904, Israel. [15]Department of Archaeology, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada. [16]Zinman Institute of Archaeology, University of Haifa, Haifa 3498838, Israel. [17]Department of Anatomy and Anthropology, Sackler Faculty of Medicine, Tel Aviv University, PO Box 39040, Tel Aviv 6997801, Israel. [18]Department of Anthropology, University of Vienna, Althanstrasse 12–14, A-1090 Vienna, Austria. [19]The Core Facility for Micro-Computed Tomography, University of Vienna, Althanstrasse 12–14, A-1090, Vienna, Austria. [20]Israel Antiquities Authority, PO Box 586, Jerusalem 91004, Israel. †Present address: Department of Evolutionary Genetics, Max-Planck-Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103, Leipzig, Germany.
*These authors contributed equally to this work.

Table 1 | Detailed dating results of the calcitic crust covering the calvaria

Sample	^{238}U (p.p.m.)	Error	Uncorr. $^{234}U/^{238}U$	Error	Corr. $^{234}U/^{238}U$	Error	Uncorr. $^{230}Th/^{234}U$	Error	Corr. $^{230}Th/^{234}U$	Error	$^{230}Th/^{232}Th$	Error	Uncorr. age (kyr)	2σ (kyr)	Corr. age (kyr)	Error 2σ (kyr)
Inner 1	3.188	0.003	1.01712	0.00162	1.02553	0.00164	0.59892	0.00139	0.40695	0.00094	3.3	0.01	99.1	1.0	56.7	9.2
Outer 2	3.753	0.003	1.01669	0.00159	1.02131	0.00160	0.52718	0.00169	0.39903	0.00128	4.4	0.01	81.2	0.8	55.4	5.2
Inner 3	2.762	0.003	1.02060	0.00100	1.03179	0.00101	0.64550	0.00162	0.45883	0.00115	3.4	0.01	112.2	1.1	66.5	10
Outer 4	3.319	0.003	1.01540	0.00113	1.02128	0.00113	0.61236	0.00282	0.46765	0.00215	4.0	0.02	102.5	1.2	68.4	6.8
Inner 6	2.771	0.007	1.01228	0.00431	1.01598	0.00433	0.46292	0.00243	0.30355	0.00159	3.6	0.01	67.3	1.1	39.2	5.6
6+7	2.555	0.003	1.01330	0.00173	1.02000	0.00174	0.58259	0.00185	0.37654	0.00119	3.2	0.01	94.8	0.3	51.1	9.4
6+7 B after US*	1.897	0.001	1.01201	0.00201	1.01785	0.00203	0.58704	0.00300	0.38947	0.00199	3.3	0.02	96.0	1.8	53.5	9.0
8A after US	2.210	0.0015	1.02594	0.00126	1.03324	0.00127	0.45589	0.00184	0.30781	0.00124	3.9	0.02	66.0	0.7	39.9	5.2
8B after US	1.343	0.0012	1.01278	0.00190	1.02083	0.00191	0.64095	0.00603	0.41945	0.00395	3.1	0.02	110.8	3.7	59.2	11.6
9A after US	1.595	0.0013	1.01574	0.00159	1.02395	0.00160	0.60011	0.00441	0.39663	0.00291	3.2	0.03	99.5	2.4	55.0	9.6
9B after US	1.423	0.0012	1.01508	0.00152	1.02428	0.00153	0.62927	0.00507	0.40843	0.00329	3.0	0.03	107.5	3.0	57.0	11.2

Analytical errors on ages are reported as 2σ; isotopic ratios are given as 1σ. Uncorr., uncorrected; Corr., corrected; Inner and Outer, the inner and outer parts of the calvaria; after US, after ultrasonic treatment (see explanation in Supplementary Information B.2).

The Manot calvaria (Manot 1) was found on a flowstone ledge within a side chamber in the northern-most area of the cave (Supplementary Information A and Extended Data Fig. 1). A thin calcite patina mixed with detrital materials, mainly clay and oxides/hydroxides, covers both its interior and exterior surfaces. The average corrected uranium–thorium (U–Th) age was obtained from 11 samples taken from different locations throughout the calcite patina (Extended Data Fig. 3). The results indicate a minimum age of 54.7 ± 5.5 kyr ago (arithmetic mean $\pm 2\sigma$); or 51.8 ± 4.5 kyr ago (weighted mean $\pm 2\sigma$) (Table 1, Supplementary Information B, Supplementary Table 1, Extended Data Table 1 and Extended Data Figs 4 and 5). The numerous speleothems inside the cave indicate that the cave was continuously wet. Thus, the calcitic crust covering the calvaria probably formed close in time to the skull's original deposition in the cave, suggesting that the minimum age closely reflects the true age of the calvaria.

Manot 1 (Fig. 2a–d) comprises the uppermost part of the frontal bone (broken 3–5 cm anterior to the coronal suture; glabella and supraciliary regions are missing), two nearly complete parietal bones and the occipital bone (broken immediately inferior to the external occipital protuberance). The calvaria is relatively small and gracile in appearance, featuring thin cranial bones (Supplementary Information C). On the basis of the synostosis of the cranial sutures, Manot 1 is the skull of an adult individual. Any determination of sex is overly speculative on the basis of the existing remains. Cranial capacity is estimated to be ~1,100 ml (Supplementary Information D).

The parietal bosses are pronounced and, in superior view, they taper gradually towards the broad frontal bone (Fig. 2a). The broadest area of the skull is high on the parietal bones, the cranial surface is flattened where the parietal bones meet along the sagittal suture, and both side walls of the skull are parallel and vertically oriented (Fig. 2c). Manot 1 thus displays features typical of modern humans (Extended Data Table 2). In contrast, in lateral aspect (Fig. 2b), the coronal suture is elevated on its external surface (coronal keel). The parietals are short, and the bregma-lambda chord and arc are small, falling outside the range of Upper Palaeolithic European and modern Mediterranean human populations (Extended Data Table 3).

The Manot occiput (Fig. 2b, d) has an occipital bun, a feature very frequently found both in European Neanderthals and in the majority of early Upper Palaeolithic modern humans. It also has a spherical suprainiac fossa (Supplementary Information C and Extended Data Table 2), which involves only the external table, leaving the internal table unaffected. The fossa is opposite the sagittal sinus bifurcation, similar to the condition seen in some modern humans, but contrasting to that usually encountered in Neanderthals. In shape, it resembles the fossae found in some of the North African Epipalaeolithic skulls of Afalou/Taforalt, and in the Upper Pleistocene skull from Aduma in Africa[7], but differs from some Levantine early AMHs, such as Qafzeh 6 or Skhul 9 (ref. 8), and from the transversely elongated fossa typical of Neanderthals[9]. The occipital plane convexity index is 22.3, similar to that of Neanderthals[10] and different from that of modern humans (Extended Data Table 3). The position of opisthocranion is well above inion, and the low placement of the transverse sulcus indicates a cerebellum that occupied less of the occipital squama than in recent modern humans (Supplementary Information C). In Manot 1, inion is located below endinion, whereas in Neanderthals inion is located superior to endinion, and in recent modern humans the two landmarks usually coincide[11]. Manot 1 has a well-developed superior nuchal line that extends across the occipital bone, similar to the morphology seen in some European Upper Palaeolithic specimens (for example, Cioclovina[12]), but unlike the morphology found in Neanderthals, in which this line is often only faintly expressed. Manot 1 lacks a horizontal occipital torus, a feature commonly seen in Neanderthals, and it fails to show a genuine external occipital protuberance. This latter feature is subtly expressed as a wide, triangular and irregular tuberculum. There is, however, a pronounced fossa between the nuchal lines for the insertion of the semispinalis capitis muscle, as is commonly described in Neanderthals[13] (Supplementary Information C).

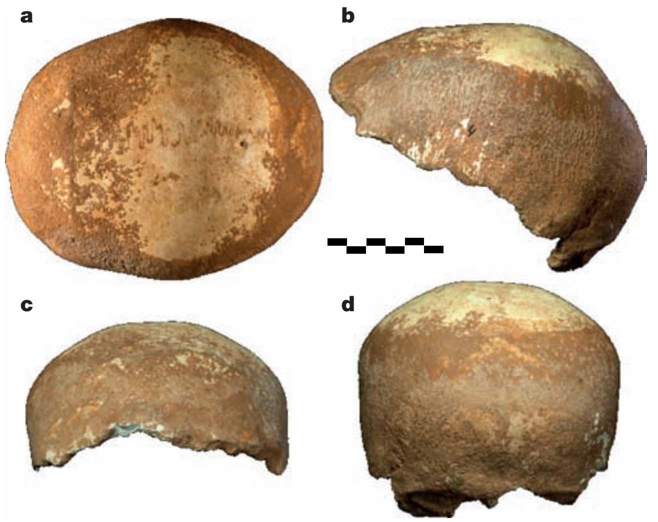


Figure 2 | Various views of the Manot 1 calvaria. **a**, Superior view. Note the coronal keel. **b**, Lateral view. Note the occipital bunning. **c**, Frontal view. Note the vertical orientation of the lateral walls, the moderate arch of the parietals towards the sagittal suture, and the flat sagittal area. **d**, Posterior view. Note the presence of a suprainiac fossa and the pronounced superior nuchal line.

On the basis of these morphological features, Manot 1 demonstrates a mosaic of 'archaic' and modern traits (Extended Data Table 2). The taxonomic significance of this combination of features is not immediately clear, but hominins with similar combinations persist in the fossil record across sub-Saharan Africa and the Levant until, and even after, ~35 kyr ago[12,14,15].

Geometric morphometric methods[16,17] were used to place the Manot 1 fossil in the broader context of the fossil record (Supplementary Information E and Extended Data Table 4). The first two principal components of shape space (Fig. 3a, b) explain ~62% of the total shape variance. This analysis places Manot 1 within the cloud of recent and Upper Palaeolithic modern humans—namely Mladeč 1 (~35 kyr cal. (calibrated years) BP), Předmostí 3 (~30–27 kyr cal. BP), Brno 2 (~29–28 kyr cal. BP), Pavlov (~29 kyr cal. BP), and Oberkassel (~14.7–13.7 kyr cal. BP)—and remote from other fossils from the Near East, such as Shanidar 1, Skhul 5 or Qafzeh 6. A nearest neighbour analysis (Fig. 3c), based on the full Procrustes distance, links the Manot specimen with recent African skulls and with central European Upper Palaeolithic specimens, such as Mladeč 1. Whereas Qafzeh 9 plots close to Manot 1 (Fig. 3a), several other Middle Palaeolithic Levantine fossils, such as Qafzeh 6 and Skhul 5, and some European Upper Palaeolithic fossils, such as Mladeč 5 and 6, are more distant. This is possibly due to a more pronounced expression of archaic traits, marked sexual dimorphism or even the existence of two different morphs at the same site[18]. Ohalo 2 (~22.8–22.3 kyr cal. BP), discovered 50 km southeast of Manot, falls close to Upper Palaeolithic humans from central Europe and to Manot 1 (Fig. 3a), but it is not among its nearest neighbours considering the full Procrustes distance (Fig. 3b).

The combination of discrete and metric data (Supplementary Information C) leads to our classification of the Manot 1 calvaria as a modern

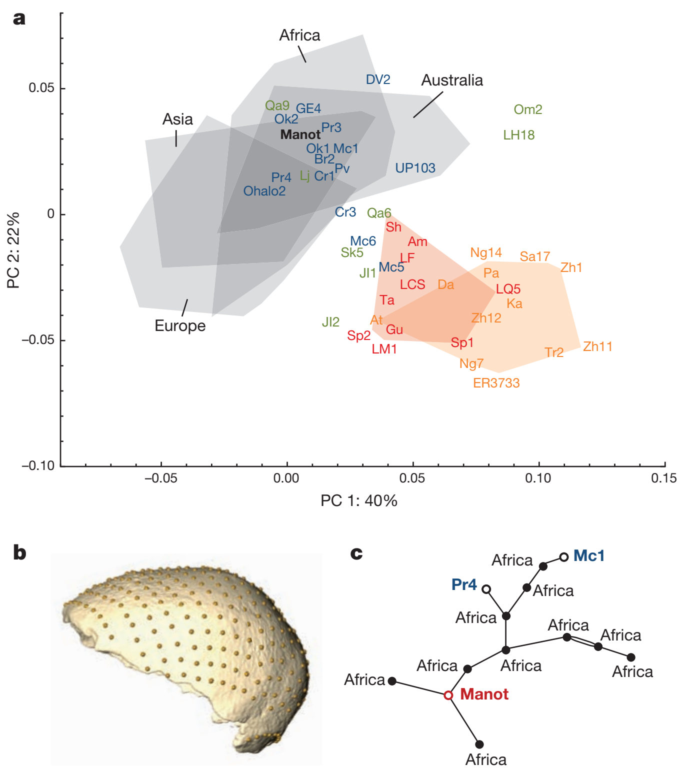


Figure 3 | Comparative morphometric analysis. **a**, First two principal components (PC) in shape space (grey, recent modern humans of diverse geographical origins; red, Neanderthals; blue, Upper Palaeolithic modern humans; green, early *Homo sapiens*; orange, archaic hominins). **b**, Landmarks and semi-landmarks used to quantify vault shape shown on a computed tomographic scan of Manot 1. **c**, Graph of the nearest neighbours based on Procrustes distance: Manot 1 is closest to recent humans from Africa, Mladeč 1 and Předmostí 4. Abbreviations appear in Extended Data Table 4.

human. Manot 1 is thus the first direct fossil evidence that modern humans inhabited the Levantine corridor ~55 kyr ago. This period coincides with the timing predicted by genetic and archaeological models for a wave of modern human dispersal out of Africa[19–22].

The Manot 1 calvaria is similar in overall shape to early Upper Palaeolithic European skulls, and its discrete features foreshadow those of later Upper Palaeolithic humans in central Europe. This implies that the Manot people could have given rise to the first modern humans to colonize Europe successfully. Thus, the anatomical features used to support the 'assimilation model' in Europe (Neanderthal–AMH interbreeding)[8] might not have been inherited from European Neanderthals but may rather have originated from earlier Levantine populations.

The possible admixture between modern humans and Neanderthals has been extensively discussed[2,3,8,23–25] (Supplementary Information C). While perspectives based primarily on morphological information generally point to Europe as the plausible location, genetic studies suggest that interbreeding occurred in a restricted geographical area, most probably in western Asia[2,3]. Despite the differences in potential localities, most predictions focus on a period later than ~100 kyr ago[2,20], most probably between 60 kyr and 50 kyr ago[26]. The close proximity, both in terms of dates and geographical location, of Manot 1 and the Levantine Neanderthals (for example, Kebara, Amud) means that the Manot 1 specimen could potentially represent a hybrid between AMHs and Neanderthals. However, any identification of potential hybrids based solely on cranial morphology must be viewed cautiously (Supplementary Information C). An earlier interbreeding, for instance, between the Tabun Neanderthals and the Skhul AMHs who may have overlapped in time[27,28], would be in disagreement with the genetic data[3,19,20,25,26].

Manot 1 could also have been a direct descendant of early AMH populations (such as Skhul/Qafzeh), but the differences in morphology between Manot 1 and the majority of fossils from these sites render this possibility unlikely (Supplementary Information C). However, it should be noted that within- and between-group morphological variations in these populations are extremely large[18], rendering any conclusion based exclusively on morphology as tentative. Nevertheless, the absence of other AMH specimens in the Levant between the Skhul/Qafzeh material (~120–90 kyr ago) and the later appearing Manot 1 (~55 kyr ago) does not support the hypothesis of continuous representation and local evolution of AMHs in the Levant.

On the other hand, the considerably fluctuating climatic conditions during MIS 5 and 4 (favouring an alteration of differently adapted populations), the unequivocal presence of Neanderthals in the region in the time gap between early AMHs and the Manot population, and the continuing evolution of AMHs in Africa[19] advocate for the most parsimonious explanation, which is that the Manot people re-colonized the Levant from Africa, rather than evolved *in situ*.

To conclude, the Manot 1 partial calvaria represents the first fossil evidence from the critical period when genetic and archaeological models predict that African modern humans successfully migrated out of Africa and colonized Eurasia. It also represents the first fossil evidence that the Levant was occupied during the late Middle Palaeolithic not only by Neanderthals (for example, Kebara/Amud), but also by modern humans. Manot 1 provides important clues about the morphology of modern humans in close chronological proximity to a probable interbreeding event with Neanderthals[3]. Our shape analysis shows that Manot 1 is a modern human, and links its morphology to recent African skulls and to some European Upper Palaeolithic fossils. This suggests that Manot 1 probably belongs to a population that had recently migrated out of Africa and established itself in the Levantine corridor during the late Middle Palaeolithic or Middle–Upper Palaeolithic interface. This time span was favourable for human migration 'out of Africa', owing to warmer and wetter climatic events over the Northern Sahara and the Mediterranean[29]. Recent evidence points to the arrival of AMHs in Europe as early as ~45 kyr cal. BP[30]. Thus, the descendants of the Manot population could have later migrated from the Levant to Europe, establishing the early Upper Palaeolithic populations there.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 May; accepted 2 December 2014.

Published online 28 January 2015.

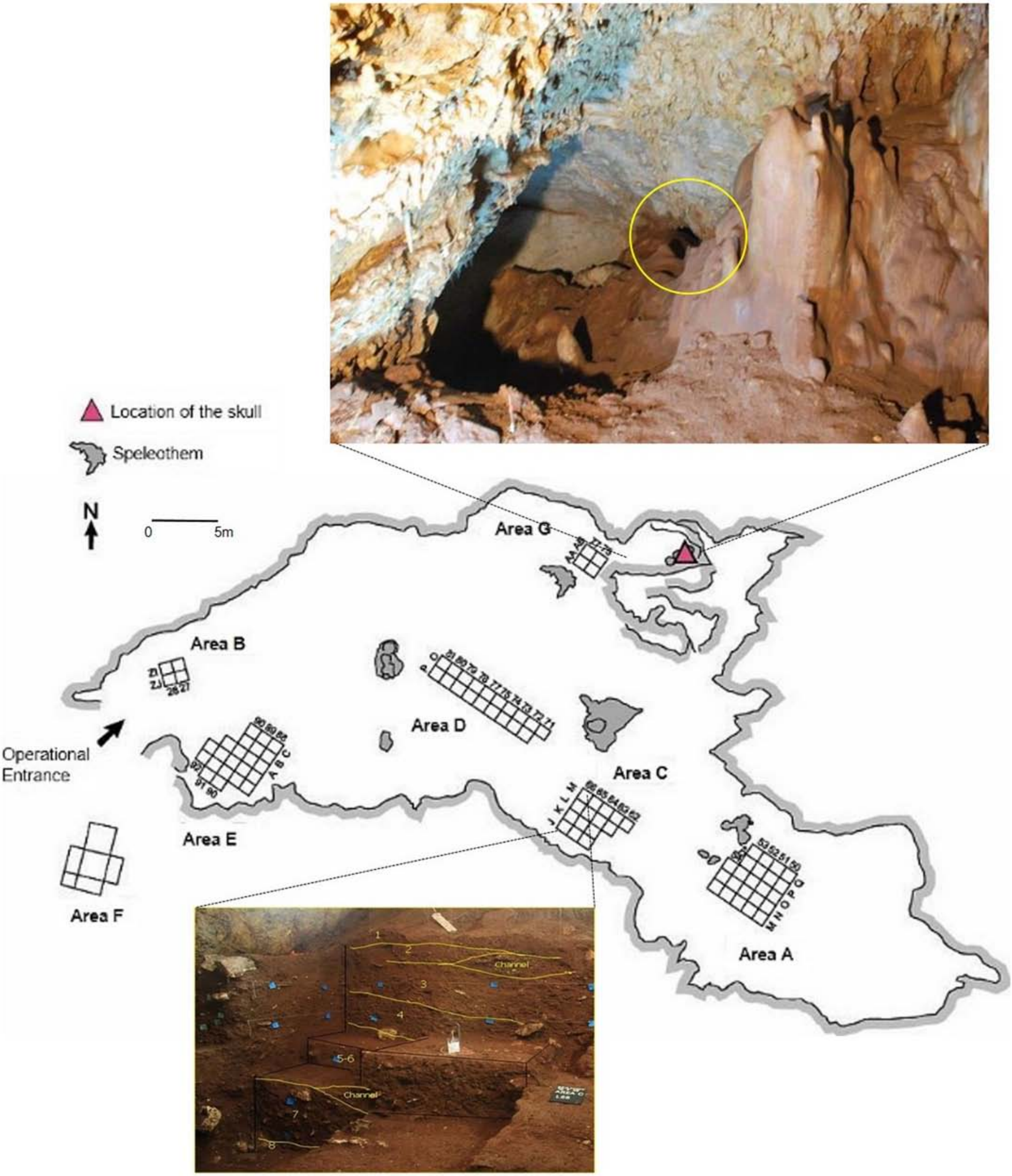
1. Hublin, J. J. The earliest modern human colonization of Europe. *Proc. Natl Acad. Sci. USA* **109,** 13471–13472 (2012).
2. Sankararaman, S., Patterson, N., Li, H., Pääbo, S. & Reich, D. The date of interbreeding between Neandertals and modern humans. *PLoS Genet.* **8,** e1002947 (2012).
3. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328,** 710–722 (2010).
4. Marder, O. *et al.* The Upper Palaeolithic of Manot Cave, Western Galilee, Israel: the 2011–12 excavations. *Antiquity* **87,** http://antiquity.ac.uk/projgall/marder337 (2013).
5. Mercier, N. *et al.* Thermoluminescence date for the Mousterian burial site of Es-Skhul, Mt. Carmel. *J. Archaeol. Sci.* **20,** 169–174 (1993).
6. Valladas, H. *et al.* Thermoluminescence dates for the Neanderthal burial site at Kebara in Israel. *Nature* **330,** 159–160 (1987).
7. Haile-Selassie, Y., Asfaw, B. & White, T. D. Hominid cranial remains from Upper Pleistocene deposits at Aduma, Middle Awash, Ethiopia. *Am. J. Phys. Anthropol.* **123,** 1–10 (2004).
8. Smith, F. H., Janković, I. & Karavanić, I. The assimilation model, modern human origins in Europe, and the extinction of Neandertals. *Quat. Int.* **137,** 7–19 (2005).
9. Balzeau, A. & Rougier, H. Is the suprainiac fossa a Neandertal autapomorphy? A complementary external and internal investigation. *J. Hum. Evol.* **58,** 1–22 (2010).
10. Hublin, J. J. in *L'Homme de Néandertal 3: l'Anatomie* 67–73 (ed. Trinkaus, E.) (E.R.A.U.L., 1988).
11. Dean, D., Hublin, J. J., Holloway, R. & Ziegler, R. On the phylogenetic position of the pre-Neandertal specimen from Reilingen, Germany. *J. Hum. Evol.* **34,** 485–508 (1998).
12. Harvati, K., Gunz, P. & Grigorescu, D. Cioclovina (Romania): affinities of an early modern European. *J. Hum. Evol.* **53,** 732–746 (2007).
13. Hublin, J. J. A propos de restes inédits du gisement de La Quina (Charente): un trait méconnu des néandertaliens et des prénéandertaliens. *Anthropologie* **84,** 81–88 (1980).
14. Crevecoeur, I., Rougier, H., Grine, F. & Froment, A. Modern human cranial diversity in the Late Pleistocene of Africa and Eurasia: evidence from Nazlet Khater, Peştera cu Oase, and Hofmeyr. *Am. J. Phys. Anthropol.* **140,** 347–358 (2009).
15. Harvati, K. *et al.* The Later Stone Age calvaria from Iwo Eleru, Nigeria: morphology and chronology. *PLoS ONE* **6,** e24024 (2011).
16. Weber, G. W. & Bookstein, F. L. *Virtual Anthropology: A Guide to a New Interdisciplinary Field* (Springer, 2011).
17. Gunz, P. *et al.* Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario. *Proc. Natl Acad. Sci. USA* **106,** 6094–6098 (2009).
18. Schwartz, J. H. & Tattersall, I. Fossil evidence for the origin of Homo sapiens. *Am. J. Phys. Anthropol.* **143,** 94–121 (2010).
19. Soares, P. *et al.* The expansion of mtDNA haplogroup L3 within and out of Africa. *Mol. Biol. Evol.* **29,** 915–927 (2012).
20. Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408,** 708–713 (2000).
21. Mellars, P., Gori, K. C., Carr, M., Soares, P. A. & Richards, M. B. Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. *Proc. Natl Acad. Sci. USA* **110,** 10699–10704 (2013).
22. Reyes-Centeno, H. *et al.* Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. *Proc. Natl Acad. Sci. USA* **111,** 7248–7253 (2014).
23. Wolpoff, M. H., Hawks, J., Frayer, D. W. & Hunley, K. Modern human ancestry at the peripheries: a test of the replacement theory. *Science* **291,** 293–297 (2001).
24. Duarte, C. *et al.* The early Upper Paleolithic human skeleton from the Abrigo do Lagar Velho (Portugal) and modern human emergence in Iberia. *Proc. Natl Acad. Sci. USA* **96,** 7604–7609 (1999).
25. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468,** 1053–1060 (2010).
26. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514,** 445–450 (2014).
27. Grün, R. *et al.* U-series and ESR analyses of bones and teeth relating to the human burials from Skhul. *J. Hum. Evol.* **49,** 316–334 (2005).
28. Grün, R. & Stringer, C. Tabun revisited: revised ESR chronology and new ESR and U-series analyses of dental material from Tabun C1. *J. Hum. Evol.* **39,** 601–612 (2000).
29. Bar-Matthews, M., Ayalon, A., Gilmour, M., Matthews, A. & Hawkesworth, C. J. Sea-land oxygen isotopic relationships from planktonic foraminifera and speleothems in the Eastern Mediterranean region and their implication for paleorainfall during interglacial intervals. *Geochim. Cosmochim. Acta* **67,** 3181–3199 (2003).
30. Higham, T. *et al.* The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* **512,** 306–309 (2014).

Supplementary Information is available in the online version of the paper.

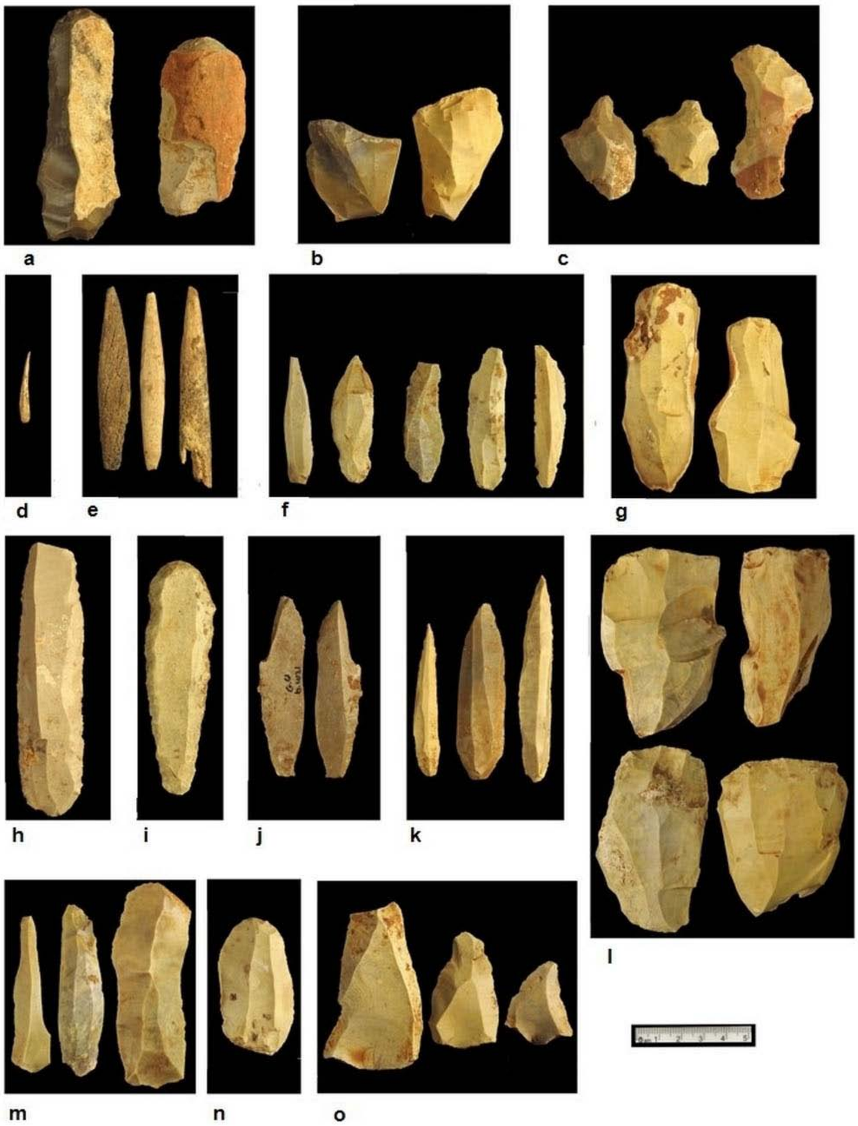
Acknowledgements The excavation at Manot Cave was initiated and supported throughout the years by the late D. David, founder of the 'Dan David Prize', and his son A. David. The ongoing research is financially supported by the Dan David Foundation, the Israel Antiquities Authority, Case Western Reserve University, the Leakey Foundation, the Irene Levi Sala CARE Archaeological Foundation, the Keren Kayemet L'Israel and the Israel Science Foundation. Radiocarbon dating research has been supported by the National Science Foundation, the Exilarch's Foundation and the Max Planck Society–Weizman Institute Joint Center for Integrative Archaeology and Anthropology. Geological research was supported by The Bertha and Louis Weinstein Research Fund. We thank other research members of the Manot team: J.-M. Tejero and S. Weiner. We thank A. Behar and L. Barda for their drawings, C. Amit for the photography of the skull, and V. Essman and Y. Shmidov for surveying and drafting the cave. We thank I. Mureinik for editorial assistance. Special thanks are due to the following students and scholars: L. Weissbrod, D. Stein, H. Cohen, B. Medlej, M. Feldman, O. Hay, T. Abulafia, L. Davis, N. Schneller-Pels, D. Yegorov, M. Ullman and G. Hertzelinger. Thanks are also due to the Maale Yossef Regional Council, the residents of modern Manot and N. Reuven. We are also grateful to the late S. Dorfman, U. Dahari, D. Barshad, E. Stern and J. Goldberg. We thank I. Gilead and O. Bar-Yosef for reading and commenting on a previous version of this paper. We are grateful to C. Stringer for comments and suggestions.

Author Contributions I.H., O.B. and O.M. are directing the Manot cave research project. I.H., P.G., B.L., V.S., G.W.W., H.M., M.G.H. and R.L.H. performed the various aspects of the anthropological study of the Manot 1 calvaria. O.B., O.M., R.L. and M.G.G. conducted the archaeological studies at the cave. A.A., M.B.-M., G.Y. and A.M. conducted the U–Th dating of the calcitic crust on the Manot 1 calvaria and of speleothems in the cave. A.F. and F.B. conducted the geological study of the cave. E.B., V.C. and B.A. performed the radiocarbon dating and charcoal analysis. G.B.-O., R.Y. and D.B.-Y.M. conducted the study of the faunal remains.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.H. (anatom2@post.tau.ac.il).

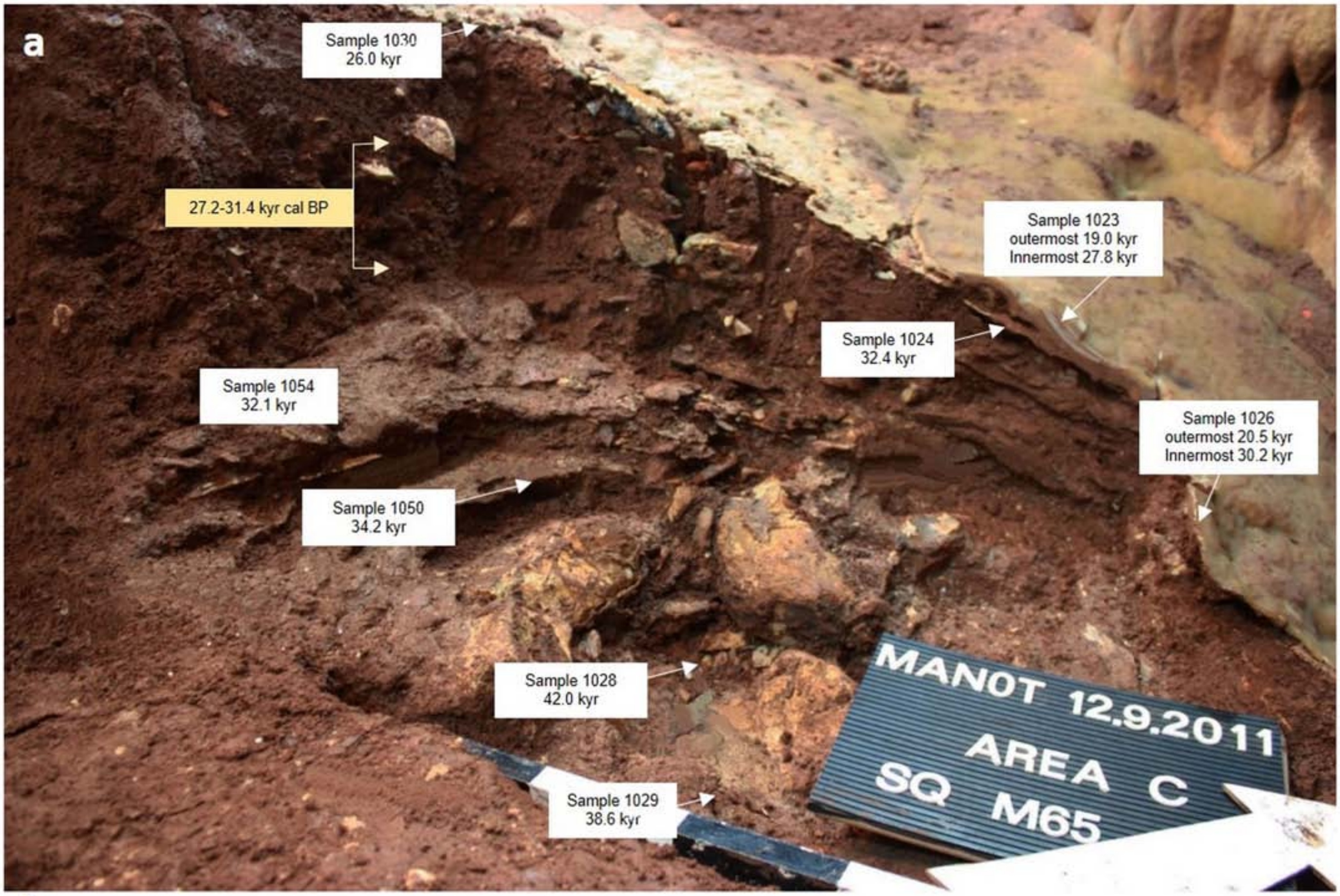


Extended Data Figure 1 | Plan of Manot Cave. The excavation areas are shown. The upper photograph shows the chamber in which Manot 1 was found (yellow circle) (looking east). The lower photograph shows the stratigraphic profile of Area C (looking west). Note the eight different sedimentological units.



Extended Data Figure 2 | Diagnostic flint and bone artefacts from Area C. Upper Palaeolithic, **a**–**k**; Initial Upper Palaeolithic, **l**–**n**; Middle Palaeolithic, **o. a**, Aurignacian endscrapers. **b**, Carinated endscrapers. **c**, Nosed endscrapers. **d**, Curved–twisted bladelet. **e**, Antler spear points. **f**, Ahmarian blades. **g**, Ahmarian blade cores. **h**, Retouched blade. **i**, Endscraper on a blade. **j**, Burin on truncation on a blade. **k**, El-Wad points. **l**, Wide blade cores. **m**, Blades with faceted butts. **n**, Endscraper. **o**, Levallois points and flakes.

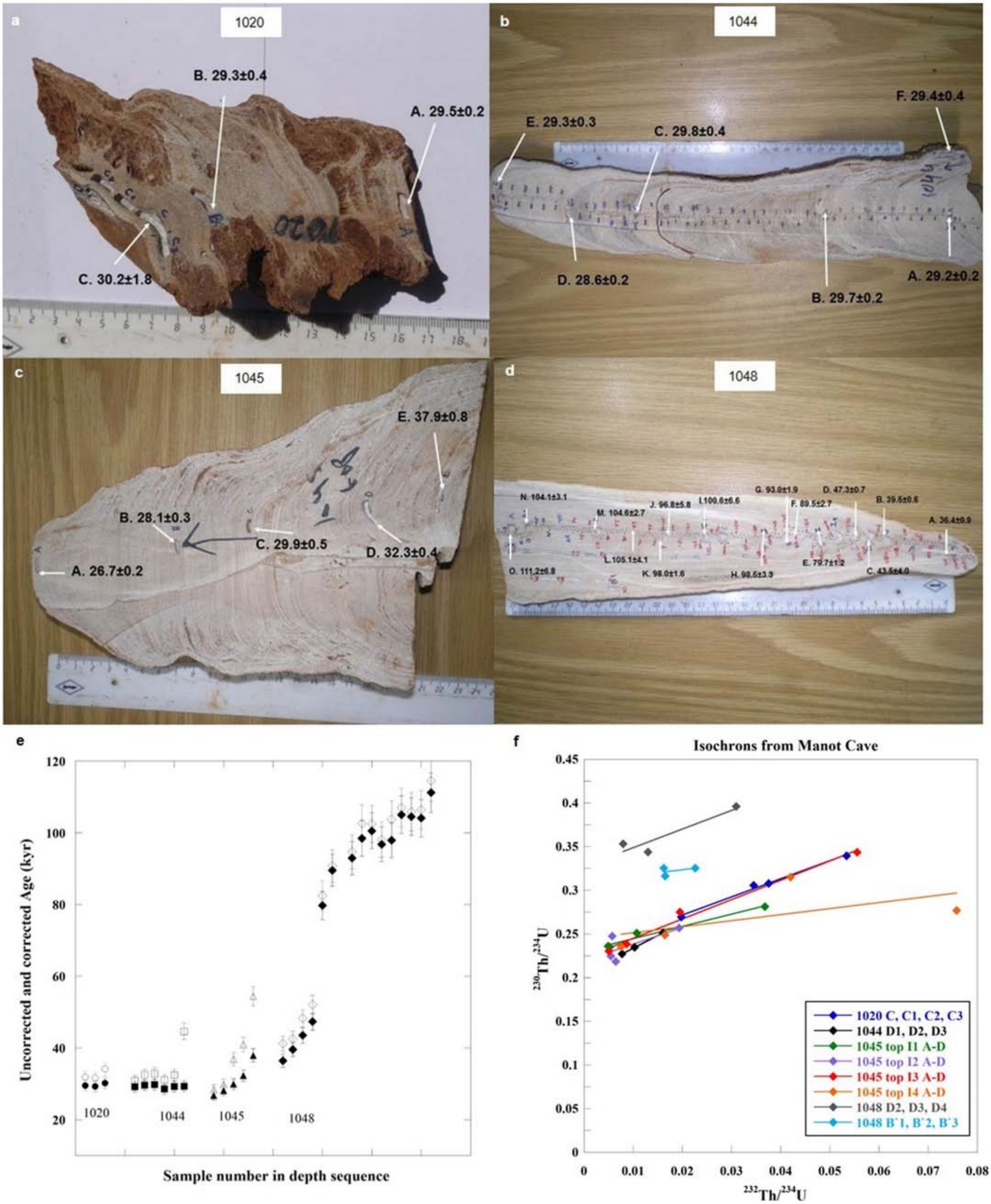
Extended Data Figure 3 | Crust sampling for dating. The outer (**a**) and inner (**b**) parts of the skull showing the locations of sampling for U–Th dating and the dating results. Ages are in thousands of years, errors are at 2σ.



b

Lab #	Type	^{14}C Age year BP	±1σ	Calibrated Age	Collection Site	Sample ID	C%	δ^{13}C PDB ‰
RTK 6704	Charcoal	23600	200	68.2% probability 27,870BP (68.2%) 27,570BP 95.4% probability 28,090BP (95.4%) 27,410BP	Area C, Sq M65a, B3607 Height: 206.23 m	C 047/12	56.6	-24.9
RTK 6705	Charcoal	23200	190	68.2% probability 27,630BP (68.2%) 27,330BP 95.4% probability 27,760BP (95.4%) 27,160BP	Area C, Sq M65a, B3613 Height: 206.14 m	C 062/12	77.5	-23.7
RTK 6706	Charcoal	26895	280	68.2% probability 31,170BP (68.2%) 30,830BP 95.4% probability 31,350BP (95.4%) 30,630BP	Area C, Sq M65a, B3619 Height: 206.10 m	C 067/12	78.3	-23.0
RTK 6708	Charcoal	24090	210	68.2% probability 28,350BP (68.2%) 27,900BP 95.4% probability 28,600BP (95.4%) 27,750BP	Area C, Sq M65a, B3624 Height: 206.05 m	C 077/12	54.8	-23.5
RTD 7116	Charcoal	48700	700	>50,000 BP Beyond calibrated range of radiocarbon	Area C, Sq J65a,b B3768 Height 205.25 m	MAN-13-347	61.1	-28.3

Extended Data Figure 4 | Dating results for Area C. **a**, U–Th dates (white boxes) are of flowstone layers and the crust covering an archaeological artefact (sample 1029). Radiocarbon dates (yellow box) are of the archaeological accumulation between the flowstone layers of square M65. **b**, Dates of radiocarbon samples from Squares M65 and J65. Radiocarbon sample number, material type, ^{14}C age $\pm$ 1σ, calibrated range for 68.2% and 95.4% confidence intervals, sample identity, percentage of carbon after pre-treatment and stable carbon isotopes ratio are given. Samples from square M65 are ordered according to absolute height above sea level.



Extended Data Figure 5 | U–Th ages of stalagmites 1020, 1044, 1045 and 1048. a-d, Photographs of the stalagmites, showing the position of the dated laminae. Ages are in thousands of years, errors are at 2σ. **e**, U–Th ages of the four stalagmites, plotted in stratigraphic sequence (open symbols, uncorrected ages; solid symbols, corrected ages). Grey and black bars indicate errors on uncorrected and corrected ages, respectively. **f**, $^{230}Th/^{234}U$ versus $^{232}Th/^{234}U$ isochron plots from eight different laminae, sampled from the four speleothems.

Extended Data Table 1 | Correction factors

	Sample number				Correction factor
Isochron	1020 C,C1,C2,C3				1.14
	1044 D1,D2,D3				0.84
	1045 I1 A,B,C,D				1.67
	1045 I2 A,B,C,D				1.24
	1045 I3 A,B,C,D				1.07
	1045 I4 A,B,C,D				3.34
	1048 D2,D3,D4				0.96
	1048 B'1,B'2,B'3				3.05
	average				**1.66± 0.35**
Age correction based on comparison with a non-corrected age from the same lamina	1023-A				1.74
	1031-C				1.74
	1044-D				1.45
	1044-D3				1.73
	1045-top-I1A				2.00
	1045-top-I1C				2.20
	1045-top-I1D				1.80
	1045-top-I2D				1.30
	1045-top-I3A				1.20
	1045-top-I3B				1.10
	1045 TOP-I3C				2.80
	1045 TOP-I4A				1.30
	1045-TOP-I4D				2.90
	1048-D4				1.60
	average				**1.78±0.15**
Wiggle matching		**Age uncorrected [kyr]**	**Age corrected [kyr]**	**Wiggle matching age [kyr]**	
	1048-A	41.3	36.4 ±0.9	37.4	2.2
	1048-C	48.3	43.5 ±4.0	46.0	3.6
	1048-D	52.1	47.3 ±0.7	48.7	2.4
	1048-H	102.6	98.5 ±3.3	96.0	0.5
	1048-I	102.5	100.6 ±6.0	96.7	0.6
	1048-K	103.8	97.9 ±1.6	97.5	1.6
	average				**1.82±0.48**

Age correction factors and the average correction factor obtained using three methods: isochrons, correction based on comparison with non-corrected ages, and wiggle-matching.

Extended Data Table 2 | Manot 1 calvaria morphology compared with an Upper Palaeolithic European specimen, Neanderthals and present-day humans

Trait	Manot 1	Mladeč 1	Neanderthals	Recent Humans
Broadest region of skull*	High on parietal, Posterior area	High on parietal, Posterior area	Low on parietal, Posterior area	High on parietal, Posterior area
Top region of calvaria	Flatter	Well rounded	Flatter	Well rounded
Parietal convexity	Shallow	Shallow	Shallow	Marked
Side wall of calvaria	Vertical	Vertical	Rounded	Vertical
Coronal suture area	Ridge	Smooth	Smooth	Smooth
Occipital bun	Present	Present	Present	Usually absent
Suprainiac fossa	Present (Rounded shape)	Absent	Present (Transverse oval shape)	Rarely present (Rounded shape)†
Occipital plane convexity‡	Marked	Marked	Marked	Shallow
Inion location	Below endinion	Below endinion	Above endinion	Varying between below and above §
Transverse sulcus‖	Cross the parietal	Cross the parietal	Cross the occipital	Cross the parietal
Superior nuchal line	Well developed	Faintly present	Variable	Variable

* The broadest region in the Manot 1 calvaria is more forward than in Mladeč 1.
† Following ref. 9.
‡ Following ref. 10.
§ Following ref. 31.
‖ In Manot 1 the most lateral part of the transverse sulcus passes through the postero-inferior angles of the parietal bone, similarly to modern humans but unlike Neanderthals and archaic humans.

31. Balzeau, A., Grimaud-Hervé, D. & Gilissen, E. Where are inion and endinion? Variations of the exo- and endocranial morphology of the occipital bone during hominin evolution. *J. Hum. Evol.* **61,** 488–502 (2011).

Extended Data Table 3 | Manot 1 calvaria measurements compared with other archaic and modern human groups

Measurement Mean ± SD (n)	**Manot 1***	Mladeč 1*	Jebel Irhoud 1†	European Neanderthals†,‡	Near East Neanderthals (Tabun, Amud, Shanidar)†	Near East AMHs (Skhul, Qafzeh)†,*	Near East Late Upper Palaeolithic (Ohalo 2)§	Near East Epipalaeolitic‖	Near East Modern‖
Max. Breadth (mm)	**135**	142	149.5	149.3±8.0 (5)	141 (Tb-1) 154 (Am-1) 154 (Sh-1)	143 (Sk-5) 145 (Sk-9) 140 (Qa-9)	144.0	M=137.1±6.3 (15) F=138.9±6.5 (8)	M=135.2±6.0 (67) F=132.5±4.7 (37)
Breg-Lambd arc (mm)	**118**	124	122	117.2±5.2 (5)	117 (Tb-1) 122 (Am-1) 131 (Sh-1)	131 (Sk-5) 120 (Sk-9) 145 (Qa-9)	126.0	M=137.4±5.0 (11) F=126.0±7.1 (16)	M=130.5±9.7 (68) F=128.1±9.0 (40)
Breg-Lambd chord (mm)	**105**	116	117	109.3±4.8 (5)	105 (Tb-1) 111 (Am-1) 118 (Sh-1)	120 (Sk-5) 112 (Sk-9) 129 (Qa-9)	116.0	M=124.6±3.9 (7) F=114.0±7.5 (4)	M=115.9±7.6 (68) F=114.1±7.5 (41)
Parietal convexity angle (degree)	**140⁰**	143.7⁰	150⁰	141.8⁰±5.4⁰ (5)	144.4⁰ (Am-1)¶	136.6⁰±1.7⁰ (3)	-	-	135.5⁰±5.5⁰ (10) M+F
Biasterion breadth (mm)	**95**	112	121	119.9±5.9 (5)	120 (Tb-1) 143 (Am-1) 118 (Sh-1)	122 (Sk-5) 120 (Sk-9) 111 (Qa-9)	107.0	M= 109.0±6.0 (3)	M= 106.9±4.8 (60) F=102.3±4.4 (33)
Occipital plane convexity (index)#	**22.3**	24.5	14.3☆	22.8±3.2 (9)**	21.1 (Am-1)¶	13.6 (Qa-9)	15.5	-	16.7±1.7 (10)†† M+F
Bone thickness at Bregma (mm)	**6.0**	7.0	7.6☆	6.7±1.4 (13)‡‡	9.0 (Am-1)¶	-	4.8	7.0±1.5 (10) M+F	6.4±1.7 (13) M+F

* Data from authors.
† Ref. 11.
‡ Including La Chapelle-aux-Saints, La Quina 5, La Ferrassie 1, Spy 1, Spy 2.
§ Ref. 32.
‖ Ref. 33 and data from authors.
¶ Ref. 34.
#Occipital plane convexity (index) was measured as the length of the perpendicular projection (subtense) from the lambda–inion chord multiplied by 100, divided by lamdba–inion chord length.
☆J. J. Hublin, personal communication.
** Ref. 10.
†† Ref. 10 measured a mean of 12.7 ± 3.5 for modern populations.
‡‡ Ref. 35. M, male; F, female; Sk, Skhul; Qa, Qafzeh; Tb, Tabun, Am, Amud.

32. Hershkovitz, I. *et al.* Ohalo II H2: A 19,000-year-old skeleton from a water-logged site at the Sea of Galilee, Israel. *Am. J. Phys. Anthropol.* **96,** 215–234 (1995).
33. Arensburg, B. *The People in the Land of Israel from the Epipaleolithic to the Present Times: A Study Based on their Skeleton Remains.* PhD thesis, Tel Aviv Univ. (1973).
34. Takai, H. & Suzuki, F. *The Amud Man and his Cave Site* 123–206 (Univ. Tokyo, 1970).
35. Lieberman, D. E. How and why humans grow thin skulls: experimental evidence for systemic cortical robusticity. *Am. J. Phys. Anthropol.* **101,** 217–236 (1996).

Extended Data Table 4 | Comparative data for the geometric morphometric analysis

Neanderthals (10)		**Archaic *Homo* (12)**	
Am	Amud 1	At*	Atapuerca SH 5
Gu	Guattari	Da	Dali
LCS	La Chapelle-aux-Saints	ER3733	KNM-ER 3733
LF	La Ferrassie 1	Ka*	Kabwe
LM	Le Moustier 1	Ng14	Ngandong 14
LQ5	La Quina H5	Ng7	Ngandong 7
Sh	Shanidar 1	Pa*	Petralona
Sp1	Spy 1	Sa17	Sangiran 17
Sp2	Spy 2	Tr2	Trinil 2
Ta	Tabun 1	Zh1	Zhoukoudian 1
Upper Palaeolithic Humans (15)		Zh11	Zhoukoudian 11
Br2	Brno 2	Zh12	Zhoukoudian 12
Cr1	Cro-Magnon 1	**Early *Homo sapiens* (8)**	
Cr3	Cro-Magnon 3	JI1	Jebel Irhoud 1
DV2	Dolní Věstonice 2	JI2	Jebel Irhoud 2
GE4	Grotte des Enfants 4	LH18*	LH 18
Mc1*	Mladeč 1	Lj	Liujiang
Mc5	Mladeč 5	Om2*	Omo 2
Mc6	Mladeč 6	Qa6	Qafzeh 6
Ohalo2	Ohalo 2	Qa9	Qafzeh 9
Ok1	Oberkassel 1	Sk5*	Skhul 5
Ok2	Oberkassel 2	**Recent Modern Humans (150)**	
Pr3	Předmostí 3	Africa	N=41
Pr4	Předmostí 4	America	N=3
Pv	Pavlov 1	Asia	N=14
UP103	Zhoukoudian Upper Cave 103	Australia	N=5
		Europe	N=85
		Papua New Guinea	N=2

The specimen abbreviations in the second column are used in the principal component analysis scores plot in Fig. 3.
* Coordinate measurements were taken on computed tomographic scans of the original.

doi:10.1038/nature14066

Distinct relationships of parietal and prefrontal cortices to evidence accumulation

Timothy D. Hanks[1,2]*, Charles D. Kopec[1,2]*, Bingni W. Brunton[1,2,3], Chunyu A. Duan[1,2], Jeffrey C. Erlich[1,2,4] & Carlos D. Brody[1,2,5]

Gradual accumulation of evidence is thought to be fundamental for decision-making, and its neural correlates have been found in several brain regions[1–8]. Here we develop a generalizable method to measure tuning curves that specify the relationship between neural responses and mentally accumulated evidence, and apply it to distinguish the encoding of decision variables in posterior parietal cortex and prefrontal cortex (frontal orienting fields, FOF). We recorded the firing rates of neurons in posterior parietal cortex and FOF from rats performing a perceptual decision-making task. Classical analyses uncovered correlates of accumulating evidence, similar to previous observations in primates and also similar across the two regions. However, tuning curve assays revealed that while the posterior parietal cortex encodes a graded value of the accumulating evidence, the FOF has a more categorical encoding that indicates, throughout the trial, the decision provisionally favoured by the evidence accumulated so far. Contrary to current views[3,5,7–9], this suggests that premotor activity in the frontal cortex does not have a role in the accumulation process, but instead has a more categorical function, such as transforming accumulated evidence into a discrete choice. To probe causally the role of FOF activity, we optogenetically silenced it during different time points of the trial. Consistent with a role in committing to a categorical choice at the end of the evidence accumulation process, but not consistent with a role during the accumulation itself, a behavioural effect was observed only when FOF silencing occurred at the end of the perceptual stimulus. Our results place important constraints on the circuit logic of brain regions involved in decision-making.

We trained rats on a previously developed decision task in which subjects accumulate sensory evidence over many hundreds of milliseconds to inform a binary left–right choice[10] ('Poisson clicks' task, Fig. 1a and Extended Data Fig. 1a–c). On each trial, rats kept their nose in a central port during the presentation of two simultaneous trains of randomly timed auditory clicks, one played from a speaker to their left and the other from a speaker to their right. At the end of the variable-duration stimulus, the rat's task was to decide which side had played the greater total number of clicks (Fig. 1a). Easy trials had a large mean rate difference between the two click trains (for example, 39:1 clicks per second), while difficult trials had a small mean rate difference (for example, 21:19 clicks per second). Accumulation of evidence models predict that averaging within a given difficulty class will produce a mean trajectory for the accumulated evidence that gradually ramps over time with a slope proportional to the mean strength of the sensory evidence (Fig. 1b). This type of correlate of evidence accumulation has been reported in several interconnected primate brain regions, including the posterior parietal cortex (PPC) and frontal eye fields (FEF)[3–5,7,8,11]. To examine whether signatures of evidence accumulation are present in the rodent brain, we recorded from 394 neurons in the PPC of 4 rats, and 397 neurons in the FOF of 6 rats while they performed the Poisson clicks task. These two areas (the PPC and FOF) have been suggested as potential rat homologues of the primate PPC and FEF, respectively[12,13]. We recorded all isolatable neurons encountered regardless of response properties. A total of 93 neurons in the PPC (23%) and 128 neurons in the FOF (32%) exhibited firing rates during the pre-movement period (from stimulus onset to centre port withdrawal) that were significantly different ($P < 0.05$) for trials that subsequently ended with a right versus left choice. This pre-movement side selectivity is consistent with previous findings in both rat PPC[14,15,19] and FOF[13]. We focus on these pre-movement side-selective neurons because they are most likely to have a role in decision formation.

Side-selective neurons in both the PPC and FOF exhibited average firing rates that were initially constant in time and then ramped upwards for stimuli in the preferred direction with a slope proportional to the stimulus strength. A similar but downwards ramping pattern was found for anti-preferred-direction stimuli (Fig. 1c, d and Extended Data Fig. 1d, e for single neuron examples). These response profiles are similar to those found in monkey PPC and FEF during visual motion discrimination tasks[3–5,11], suggesting that rats and monkeys may use similar algorithms and related circuits to perform evidence accumulation.

We extracted the average influence of a single quantum of sensory evidence (a click) on the neural firing rates by computing the click-triggered average firing rate. Individual clicks have a measurable and sustained influence on PPC neural responses (Fig. 1e), consistent with the sustained effects of analogous manipulations in monkey PPC[16]. Individual clicks also produced a sustained influence on FOF neural responses, albeit with a magnitude that slowly but significantly decayed over several hundred milliseconds (Fig. 1f). No equivalent experiment from primate FEF exists as yet for comparison. The longer mean latency of the click-triggered average in the PPC (~200 ms) compared with the FOF (~100 ms) suggests that sensory information in this task may reach the FOF through circuits that bypass the PPC.

We used the behavioural model previously developed for this task[10] to obtain trial-by-trial, moment-by-moment estimates of the accumulating evidence, denoted by $a(t)$ (Extended Data Fig. 2 and Extended Data Table 1). The model provides the formal framework with which to combine all available behavioural data for estimating $a(t)$. Click times and rat choices from all trials determine the model's parameters. Click times and the rat's choice on individual trials determine the inferred evolution of the accumulator for the corresponding trial. Together with the simultaneously recorded firing rates $r(t)$, this enabled the estimation of 'tuning curves' that specify, for each point in time during the stimulus, how r depends on a (Fig. 2).

This analysis revealed that, on average across the population of neurons, the PPC encoded the accumulator's evolving value $a(t)$ using a graded map that is stable across time (Fig. 3a). A similar encoding was observed in monkey PPC, albeit using a task that did not require a temporal accumulation strategy[17]. Thus, during decision formation, firing rates in the PPC change over time as the accumulated evidence changes, but at any time point the graded value of the accumulator is encoded, by a fixed map, into a graded firing rate. In this way, the firing rate encodes

[1]Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey 08544, USA. [2]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. [3]Departments of Biology and Applied Mathematics, University of Washington, Seattle, Washington 98105, USA. [4]NYU-ECNU Institute of Brain and Cognitive Science, NYU-Shanghai, Shanghai 200122, China. [5]Howard Hughes Medical Institute, Princeton University, Princeton, New Jersey 08544, USA.
*These authors contributed equally to this work.

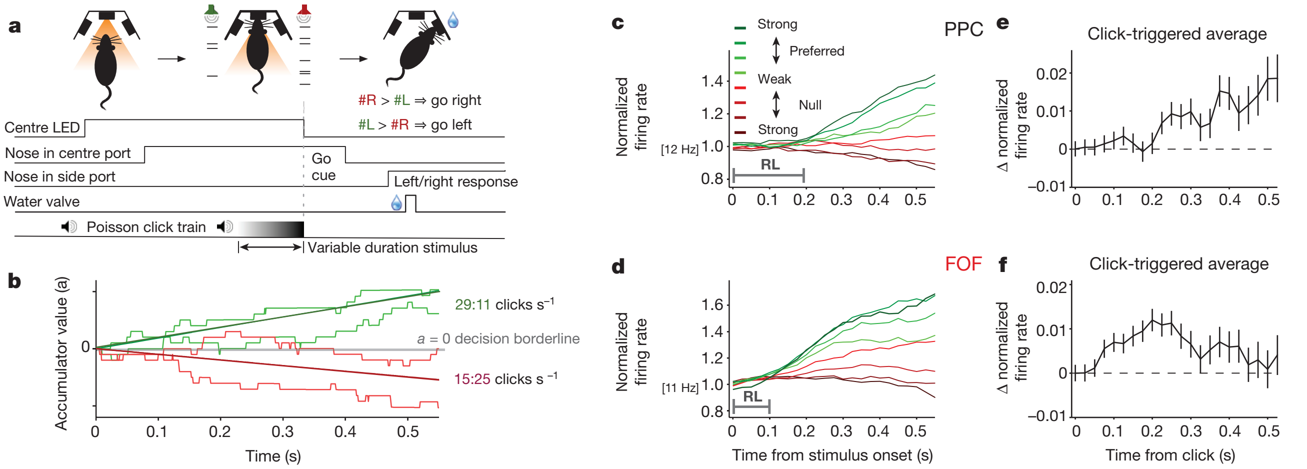


Figure 1 | Side-selective neurons in PPC and FOF exhibit signatures of evidence accumulation. **a**, Sequence of events for each trial. **b**, Schematic of the evidence accumulation process. Each right (left) click provides a single quantum of positive (negative) evidence. The thick green line shows the expected trajectory, averaged over many trials, of the accumulating evidence a for a mean stimulus strength of 29:11 clicks s^{-1}. The red line shows this for 15:25 clicks s^{-1}. The lighter-coloured traces show how individual trials within a given stimulus-difficulty class meander based on the variable click times of each trial. At the 'go' signal (offset of centre light-emitting diode (LED)), the sign of a indicates the appropriate decision. **c**, PPC average population responses. Trials were grouped by average strength of the sensory evidence. Green traces correspond to stimuli in the preferred direction of the recorded neurons, and red traces to the non-preferred direction. Darker hues correspond to easier trials. Responses exhibit ramping profiles that depend on the mean stimulus strength. The PPC response lag (RL) until stimulus-strength-dependent ramping was ~200 ms. $n = 93$ neurons from 4 rats. **d**, Same as **c** for neurons in the FOF. Similar to PPC, responses exhibit ramping profiles that depend on the mean stimulus strength. Response lag in the FOF is ~100 ms. $n = 128$ neurons from 6 rats. **e**, Click-triggered average response ± s.e.m. in the PPC. **f**, Same as **e** for FOF.

the answer to the question 'what is the value of the mentally accumulated evidence?'.

Despite the similarities between the FOF and PPC based on trial-averaged responses (Fig. 1c, d), the tuning curve analysis revealed a very different encoding of the value of the accumulator $a(t)$ in the FOF. Neurons in the FOF exhibited a more categorical mapping between firing rate and the sign of the accumulator value that tended to cluster into an $a > 0$ and an $a < 0$ group, while the mapping was again stable across time (Fig. 3b). The sign of a indicates the choice, so the FOF categorically encodes which of the two available choices is favoured by the evidence accumulated so far. Thus, FOF responses can be approximately described as the categorical answer to the question 'if the go signal came now, which choice should I make?'.

A direct comparison of the time-averaged tuning curves showed a significantly larger slope at the zero-crossing for the FOF compared to the PPC (Fig. 3c; PPC: 0.058 ± 0.003, FOF: 0.158 ± 0.015, mean ± 95%

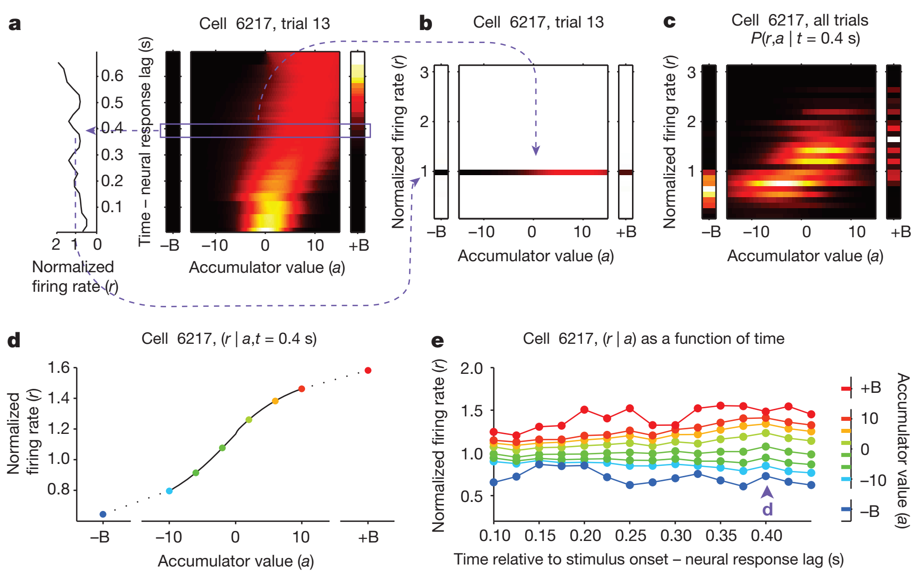


Figure 2 | Computing tuning curves that describe the relationship between neural activity and accumulated evidence. **a**, One trial for an example neuron from the PPC. The left side shows the firing rate of the neuron, and the right side shows the behavioural model's estimate of the evolution of the distribution of a (colour represents probability density). Time runs vertically and is aligned to stimulus onset minus neural response lag. ±B correspond to the 'sticky' decision-commitment bounds on evidence accumulation. **b**, Building a map of firing rate versus accumulator value. At a given time point (here, $t = 0.4$ s), we copy the distribution of a (purple box) to a vertical position given by the firing rate of the neuron. **c**, Continuing with the same time point, we add a slice from every recorded trial. This produces the full joint distribution $P(r,a \mid t = 0.4)$, the probability of seeing firing rate r and accumulator value a at time $t = 0.4$ s. **d**, The accumulator values are binned, and the mean firing rate is computed for each bin to generate a neural tuning curve as a function of a. **e**, The process is repeated for each time point. Each vertical slice corresponds to a tuning curve, with the one from **d** shown above the purple arrowhead.

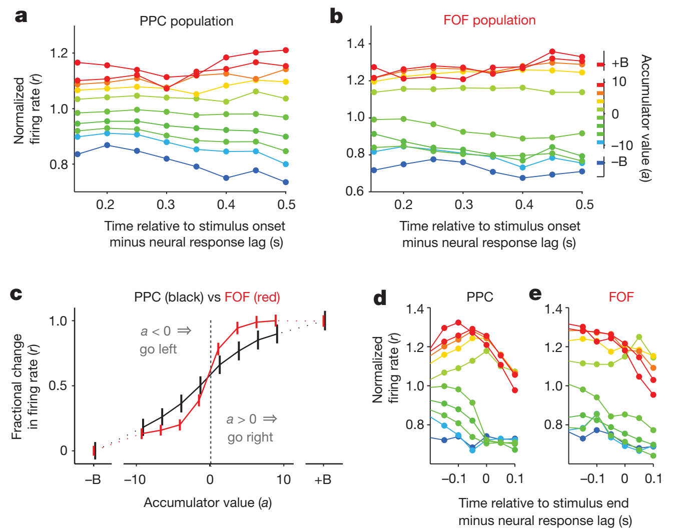


Figure 3 | PPC encodes graded accumulated evidence while FOF has a more categorical encoding. **a**, Relationship between the PPC population firing rate and accumulator value aligned to the stimulus onset minus the neural response lag (200 ms for PPC). Colours indicate accumulator value with ±B corresponding to sticky accumulation bounds. The $a \to r$ map in PPC is graded and relatively stable over time. $n = 93$ neurons. **b**, Same as **a** for the FOF (100 ms neural response lag). The $a \to r$ map is again relatively stable over time, but the responses appear to be more clustered based on the sign of a. $n = 128$ neurons. **c**, Average ± s.e.m. over the full time period shown ($t = 0.15$ s to $t = 0.5$ s). PPC, in black; FOF, in red. Responses were scaled to span the range from 0 to 1 to account for differences in dynamic range. PPC shows a smoothly graded relationship, while FOF shows a sharper dependence on the sign of the accumulator value. **d**, Relationship between PPC population and accumulator value aligned to the putative end of the decision process, which is defined as the time relative to the stimulus end minus the response lag.

confidence interval), which indicates a sharper transition of firing rate between negative and positive accumulator values for the FOF. This difference was robust to variations in the time constant of integration and the value of a 'sticky' decision commitment bound in the behavioural model of evidence accumulation, variations in neural response latencies, and whether the behavioural model was fit individually to each rat or to the aggregate behaviour for all rats (Extended Data Figs 3 and 4).

The difference between the PPC and FOF was also apparent at the level of individual neurons. We computed the steepness of the tuning curve of each neuron, and while there was overlap between the PPC and FOF populations, there was a significant shift towards greater steepness for the distribution of FOF neurons compared to PPC neurons ($P < 0.001$, Extended Data Fig. 5). This difference in encoding between the PPC and FOF was consistent both with the response profiles (Fig. 1c, d) and the click-triggered average responses, including the diminishing trend for the FOF (Extended Data Fig. 6). The difference in encoding between the PPC and FOF was not maintained after the end of the decision process. During the period of motor preparation, the FOF maintained a categorical encoding, while the graded encoding in the PPC converged to a categorical encoding (Fig. 3d, e).

The more categorical encoding found in the FOF suggests that, contrary to current views[3,5,7–9], this brain region may not be involved in the graded evidence accumulation process itself, but instead may be more involved in the conversion to a categorical choice. Such a view predicts that FOF activity should have an effect on behaviour only at the end of stimulus presentation, which is when information about the categorical choice will be used to generate a motor response. By contrast, if the FOF were involved in evidence accumulation—a process that occurs throughout the stimulus presentation[10]—its activity should have an effect on decision formation throughout the entire sensory stimulus.

To test these predictions, we used the inhibitory opsin halorhodopsin eNpHR3.0 to inactivate the FOF unilaterally and transiently during the Poisson clicks task (Extended Data Fig. 7). Full-trial inactivation (2-s period from 500 ms before auditory stimulus onset until 500 ms after stimulus end, Fig. 4a) resulted in a significant ipsilateral choice bias (10.3 ± 3.0%, $P < 0.01$, mean ± s.e.m. across rats). We next assessed the temporal specificity of the effect of FOF inactivation using four different 500-ms time periods: the delay before stimulus onset, the first half of a 1-s stimulus, the second half of a 1-s stimulus ('peri-choice'), or the movement period ('post-choice'). Only peri-choice inactivation led to a significant ipsilateral bias (Fig. 4b, 10.6 ± 1.0%, $P < 0.01$). Inactivation during the early accumulation period produced a smaller effect ($P < 0.01$) that was not significantly different from zero ($P = 0.48$). In a second group of rats we used even shorter inactivation periods: either the next-to-last, or the final (peri-choice) 250 ms of a variable-duration click train. Again, only the peri-choice perturbation had an effect on behaviour (Fig. 4c, 5.4 ± 0.8%, $P < 0.01$), while the effect of perturbation just 250 ms earlier was smaller than the peri-choice effect ($P < 0.01$) and not statistically significant ($P = 0.45$). Furthermore, in both the

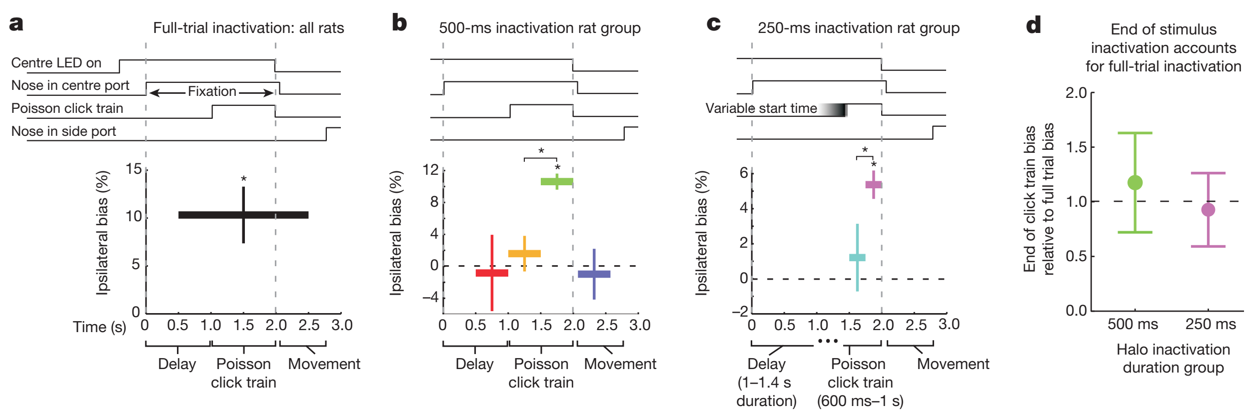


Figure 4 | Temporally precise transient halorhodopsin (eNpHR3.0)-mediated inactivation reveals that FOF activity has a significant effect on decision formation only at the end of the auditory stimulus presentation. **a**–**c**, The task structure is shown at the top. For the bottom panels, the horizontal extent of the thick black or coloured bars indicates the period of inactivation (laser on). The vertical position indicates the average bias across rats induced by the corresponding period of FOF inactivation (percentage 'went ipsi' laser on – percentage 'went ipsi' laser off). *$P < 0.01$ (bootstrap) of effect size or comparison. **a**, Bias caused by 'full-trial' (2 s long laser pulse) inactivation. $n = 18$ rats. **b**, Bias caused by 500-ms inactivation during the pre-stimulus period (red; $n = 5$ rats), the first half of the stimulus (yellow; $n = 9$), the second half of the stimulus ('peri-choice', green; $n = 9$), or the post-stimulus movement (purple; $n = 5$). **c**, Bias caused by 250-ms inactivation during the next-to-last 250 ms (light blue; $n = 7$), or the final 250 ms (peri-choice, magenta; $n = 7$) of a 0.6–1.0 s duration stimulus. **d**, Peri-choice inactivation bias normalized by full-trial inactivation bias (500 ms group in green; 250 ms, in magenta). Normalization performed for each rat independently. For all panels, error bars show s.e.m. across rats.

500-ms and 250-ms groups of rats, the magnitude of the bias induced by peri-choice inactivation fully explained the magnitude of the bias for full-trial inactivation (Fig. 4d, peri-choice normalized by full-trial bias = 1.17 ± 0.45 and 0.93 ± 0.33 for each group, respectively). Consistent with the idea that the dominant role of the FOF is to control the categorical choice, a model-based analysis indicated that a post-categorization bias explained these optogenetic inactivation data significantly better than alternative forms of bias that directly affected the accumulation process (J.C.E. *et al.*, manuscript in preparation; $P < 0.05$, see Methods; Extended Data Fig. 8). Finally, and again consistent with the FOF having a role that is separate from the click accumulation process, we found no correlation between choice biases induced by unilateral perturbation and click counts or stimulus duration (Extended Data Fig. 9).

Our results indicate that rather than participating in a single, distributed process during decision-making[18], neurons in parietal and prefrontal areas have distinct relationships to accumulating evidence. PPC neurons veridically encode a graded value of the accumulated evidence, even though separate work from our laboratory suggests that PPC activity is not necessary for choice behaviour in this task (J.C.E., manuscript in preparation). By contrast, perhaps in readiness for the 'go' signal that comes at a variable duration after stimulus onset, neural activity in the FOF can be approximately described as representing, throughout the stimulus, the categorical answer to the question 'if the go signal came now, which choice should I make?'. Unilateral FOF perturbations affect choice behaviour only at the end of the evidence accumulation period, when the provisional choice must be converted into a motor act. These results suggest that accumulation occurs upstream to FOF and challenge the prevailing view that the prefrontal cortex is part of the neural circuit for accumulation of evidence[3,5,7,8] (see Supplementary Discussion). The FOF may instead be necessary for the final step in the decision process: the conversion of the graded accumulation signal into a categorical choice.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 March; accepted 12 November 2014.
Published online 19 January 2015.

1. Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30,** 535–574 (2007).
2. Ding, L. & Gold, J. I. Caudate encodes multiple computations for perceptual decisions. *J. Neurosci.* **30,** 15747–15759 (2010).
3. Ding, L. & Gold, J. I. Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cereb. Cortex* **22,** 1052–1067 (2012).
4. Shadlen, M. N. & Newsome, W. T. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* **86,** 1916–1936 (2001).
5. Kim, J. N. & Shadlen, M. N. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neurosci.* **2,** 176–185 (1999).
6. Ratcliff, R., Hasegawa, Y. T., Hasegawa, R. P., Smith, P. L. & Segraves, M. A. Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *J. Neurophysiol.* **97,** 1756–1774 (2007).
7. Purcell, B. A. *et al.* Neurally constrained modeling of perceptual decision making. *Psychol. Rev.* **117,** 1113–1143 (2010).
8. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503,** 78–84 (2013).
9. Heitz, R. P. & Schall, J. D. Neural mechanisms of speed-accuracy tradeoff. *Neuron* **76,** 616–628 (2012).
10. Brunton, B. W., Botvinick, M. M. & Brody, C. D. Rats and humans can optimally accumulate evidence for decision-making. *Science* **340,** 95–98 (2013).
11. Roitman, J. D. & Shadlen, M. N. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* **22,** 9475–9489 (2002).
12. Whitlock, J. R., Sutherland, R. J., Witter, M. P., Moser, M.-B. & Moser, E. I. Navigating from hippocampus to parietal cortex. *Proc. Natl Acad. Sci. USA* **105,** 14755–14762 (2008).
13. Erlich, J. C., Bialek, M. & Brody, C. D. A cortical substrate for memory-guided orienting in the rat. *Neuron* **72,** 330–343 (2011).
14. McNaughton, B. L. *et al.* Cortical representation of motion during unrestrained spatial navigation in the rat. *Cereb. Cortex* **4,** 27–39 (1994).
15. Whitlock, J. R., Pfuhl, G., Dagslott, N., Moser, M.-B. & Moser, E. I. Functional split between parietal and entorhinal cortices in the rat. *Neuron* **73,** 789–802 (2012).
16. Huk, A. C. & Shadlen, M. N. Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J. Neurosci.* **25,** 10420–10436 (2005).
17. Yang, T. & Shadlen, M. N. Probabilistic reasoning by neurons. *Nature* **447,** 1075–1080 (2007).
18. Andersen, R. A. & Cui, H. Intention, action planning, and decision making in parietal-frontal circuits. *Neuron* **63,** 568–583 (2009).
19. Raposo, D., Kaufman, M. T. & Churchland, A. K. A category-free neural population supports evolving demands during decision-making. *Nature Neurosci.* **17,** 1784–1792 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Deisseroth for support with optogenetics. We thank A. Akrami, T. Buschman, J. Gold, B. Pesaran, B. Scott, D. Tank and M. Yartsev for comments on the manuscript. We thank A. Begelfer, K. Osorio and J. Teran for animal and laboratory support. T.D.H. was supported by National Institutes of Health (NIH) Award Number F32MH098572. C.A.D. was supported by a Howard Hughes Medical Institute predoctoral fellowship. C.D.K. was supported in part by the NIH Award Number T32MH065214.

Author Contributions T.D.H., B.W.B., C.A.D. and J.C.E. collected electrophysiological data. T.D.H. analysed electrophysiological data. J.C.E. played an advisory role on electrophysiological experiments. C.D.K. carried out the optogenetic experiments, with assistance from B.W.B. C.D.K. analysed the optogenetics data, with input and assistance from T.D.H. and J.C.E. T.D.H., C.D.K. and C.D.B. wrote the paper. C.D.B. was involved in all aspects of experimental design and data analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.D.B. (brody@princeton.edu).

METHODS

Subjects. A total of 27 male Long–Evans rats (*Rattus norvegicus*) between the ages of 6 and 24 months were used for this study. Of these, 9 rats were used for neural recordings (3 for PPC alone, 5 for FOF alone, and 1 for PPC and FOF combined), and 18 for optogenetic inactivation. All statistical tests were made between groups with similar sample sizes. Investigators were not blinded to experimental groups during data collection or analysis. Animal use procedures were approved by the Princeton University Institutional Animal Care and Use Committee and carried out in accordance with National Institutes of Health standards.

Behaviour. Rats were trained to perform an auditory evidence-accumulation decision task[10]. For each session, rats were placed in a behavioural training box. At the start of each trial, rats are instructed by an LED to place their nose in a central port and maintain nose fixation. After a delay of at least 0.2 s, they were presented with a train of randomly timed clicks from a left speaker and, simultaneously, a different train of clicks from a right speaker. The click trains were generated by Poisson processes with different underlying rates. For neural recording sessions, the duration of stimulus presentation was varied randomly from trial to trial, with durations ranging from 0.1 to 1.2 s. For full-trial and 500-ms optogenetic inactivation sessions, the stimulus duration was 1 s for all trials. For the 250-ms optogenetic inactivation sessions, the stimulus duration ranged from 0.6 to 1.0 s. When the stimulus ended, the central LED was turned off as a 'go' signal, and rats indicated a choice by making an orienting movement to one of the side nose ports. They were rewarded with a water drop from that port if they chose the side that played more clicks. Trial difficulty was also manipulated by changing the ratio of right to left clicks, with the total mean click rate held fixed at 40 Hz. For neural recording sessions, mean click rate ratios varied from 39:1 clicks s^{-1} for the easiest trials to 20:20 clicks s^{-1} for the most difficult ones. For optogenetic inactivation sessions, the mean click rate ratios varied from 39:1 clicks s^{-1} to 26:14 clicks s^{-1}. Model-based (Extended Data Table 1) as well as model-free (Extended Data Fig. 1) analyses were used to confirm that the rats were basing their decisions on evidence gradually accumulated throughout each stimulus period.

Recordings. Rats were implanted with custom-made microdrives and recordings were made with platinum-iridium tetrodes, as described previously[13]. FOF recording sites were centred at +2 anterior–posterior (AP), ±1.3 medio–lateral (ML) (mm from Bregma); PPC sites were centred at −3.8 to −4.1 AP and ± 2.2 ML. The choice of recording area in each particular rat was assigned randomly. Electrode placements were confirmed with histology.

Neural data analysis. Action potentials were isolated as belonging to single neurons by performing manual cluster cutting of tetrode waveforms from events that exceeded a threshold level on any of the tetrode channels. We recorded spike times for all isolatable neurons encountered regardless of response properties. We then selected for further analysis those neurons that had a mean response at stimulus onset larger than 1 Hz and significantly different firing rates during the pre-movement period for trials that subsequently end in right versus left choices ($P < 0.05$, Receiver operating characteristic). The 'pre-movement period' was defined to start at stimulus onset and end at response initiation. We defined the choice side that yielded the larger response as the preferred side. The preferred side was contralateral for 59.1% of the PPC neurons and 59.4% of the FOF neurons. We note that mixed selectivity like this in rat prefrontal cortex still yields ipsilateral biases after unilateral inactivation[13,20].

Individual trial rate functions were generated by smoothing the spike trains with a causal half-Gaussian filter with 0.1 s standard deviation. Before combining responses into a population average, responses were normalized by the mean response for each neuron at stimulus onset. To generate peri-event time histograms (PETHs), responses were sorted for each cell by the mean stimulus strength. Mean stimulus strength was defined by dividing trials for each cell into 8 quantiles based on preferred minus non-preferred click rates—that is, the difference in the number of clicks presented divided by the stimulus duration. More precisely, positive and negative click rate trials (those with more preferred and non-preferred clicks, respectively) were separately sorted into 4 quantiles each. Trials with the same number of preferred and non-preferred clicks were assigned in equal numbers to the two quantile bins with lowest absolute magnitude rates—that is, those closest to zero. For all analyses aligned to stimulus onset, trials were included only up to the time of movement onset.

To measure the effects of single clicks on neural responses, 'click-triggered averages' were calculated as follows. For each neuron, we first grouped the trials based on the underlying Poisson rates used to generate the stimuli. For each group, we calculated the mean PETH, which corresponds to the expected neural response at each point in time for each Poisson rate group. We subtracted this mean response from each member (trial) in the group to yield the residual response with respect to the expectation given the Poisson rate. Owing to the Poisson nature of the stimulus, given an underlying click-generating Poisson rate, the presence of a click at one moment in time has no bearing on the probability of a click at other times. Therefore, if one aligns the residual response to a click, it describes the change in response associated with that click relative to the average expected response to clicks at other times. In this way, the residual response allowed us to isolate the effect of a single click apart from the confounding effect of the mean response across time for a given Poisson rate. Thus, we aligned the residual responses to the time of each click and averaged across these alignments to obtain the click-triggered average response across all clicks for each Poisson rate group; the click-triggered average for each neuron was the result of averaging across groups. To combine across both preferred-direction and non-preferred-direction clicks, the residual responses for non-preferred-direction clicks were inverted in sign before averaging.

To make model-based predictions for the click-triggered average, we used the model to simulate evidence accumulation trajectories for 5,000 trials based on the same range of stimulus difficulties and durations used to collect neural data. We then encoded these trajectories with either a graded, linear scheme (veridical encoding of accumulated evidence, firing rate $r = k_1 \times a(t) + k_2$, in which k_1 and k_2 are constants) or a categorical scheme (firing rate $r = k_1 \times \mathrm{sign}(a(t)) + k_2$, in which k_1 and k_2 are constants). For the graded scheme, we introduced an encoding lag with a mean of 0.2 s to match the response lag of the PPC. For the categorical scheme, we introduced an encoding lag with a mean of 0.1 s to match the response lag of the FOF. For both schemes, the encoding lag was taken from a Gaussian distribution with 0.02 s standard deviation to account for variability in lag across neurons. We then used the same methods described above to measure the click-triggered average of these two encodings.

We used a similar approach to calculate the expected PETH response profiles that would be obtained for a categorical-encoding model (Extended Data Fig. 6). To do this, we simply averaged responses from the categorical-encoding model across trials, sorting based on mean stimulus strength with the same method as described above for Fig. 1.

Behavioural model-based analysis of neural data. To generate tuning curves that specify the relationship between neural firing rates and mentally accumulated evidence, we exploited a trial-by-trial behavioural model of the evidence accumulation decision process[10] (Extended Data Fig. 2). The model converts the incoming stream of each trial of discrete left and right click stimuli into a scalar quantity $a(t)$ that represents the gradually accumulating difference between the two streams and drives choices: at the end of the stimulus, if a is positive (negative), the model prescribes 'choose right' ('choose left'). The rat's behaviour is used to fit parameters that govern how $a(t)$ evolves. These parameters quantify sensory and accumulator noise, leakiness/instability of the accumulation process, a sticky accumulation bound, sensory depression/facilitation, side bias, and a lapse rate that corresponds to a fraction of trials on which a random choice is made. Quantifying and taking into account the noise sources implies that on each trial, the model estimates the evolution of a noise-induced probability distribution over values of the scalar $a(t)$. Note that in addition to the probability distribution over $a(t)$, each trial also has an associated lapse probability, which is constant as a function of time. At any moment of time, the forward version of the model allows for the estimation of the distribution of the accumulated evidence given the stimulus that was presented. A better estimate of the accumulated evidence can be obtained by also taking into account the choice made at the end of the trial. This is calculated using the backward propagation of accumulator values that are consistent with that choice at the end of the stimulus presentation[10] (Extended Data Fig. 2b). We use this improved estimate that also takes into account subject's choice to relate the accumulator value to neural responses.

For the analyses presented in the main text, the behavioural model was fit separately to each rat (see model parameters, Extended Data Table 1). Thus, in estimating the value of $a(t)$, variability between individual rats was taken into account. To ensure that this variability did not itself result in important effect differences, we also performed separate analyses in which we estimated $a(t)$ from a single set of model parameters derived from the combined behaviour across all rats (Extended Data Fig. 4).

In addition to these estimates of $a(t)$, we also have an estimate of the response of each neuron as a function of time from the single trial rate functions described above. The estimates of the neural response and accumulated evidence were then used to calculate the joint probability distribution between those two variables as a function of time for each neuron. The correspondence between time in the model and neural time was determined based on the latency of the stimulus-dependent responses in each region. Thus, $t = 0$ in the model was taken as 200 ms after stimulus onset for the PPC, and 100 ms for the FOF. From the joint probability, we extracted the response of the neuron given the value of the accumulator. We combined across neurons by weighting the contribution of each by the inverse of the variance of this conditional distribution. This gives more weight to representations that are less noisy.

To quantify the relationship between neural response and accumulator value, we first averaged across the time period from 0.15 to 0.5 s into the decision process (Fig. 3c). To account for the variety of dynamic ranges of the neural responses, we

scaled the responses to span the range from 0 to 1, either for individual neurons (Extended Data Fig. 5a) or for each cortical region (Fig. 3). We then fit the relationship of the response to the accumulator with a four-parameter sigmoid with the following equation:

$$r = k_1 + \frac{k_2}{1 + e^{-k_3(a - k_4)}}$$

In this equation, $k_2 k_3/4$ determines the slope at zero-crossing, which characterizes whether the neural response changes smoothly between negative and positive accumulator values or whether it changes sharply in this region. We extracted this parameter and its confidence interval for both the PPC and FOF populations. Statistical comparisons between the populations were calculated using nonparametric bootstrap procedures resampling from the populations of neurons with replacement. We also extracted the shape parameter for individual neurons and used the nonparametric Mann–Whitney U (rank sum) test to determine whether the medians of each population distribution were significantly different.

We also performed several analyses to ascertain to what extent specific parameter values were critical for our results. We focused on the time constant of accumulation and the sticky decision bounds, because these two parameters describe the extent to which psychophysical accumulation differs from optimality. For both parameters, we scaled its best-fit value by factors of 0.5, 0.67, 1.5 and 2, and compared the slopes of the tuning curves for PPC and FOF based on the revised estimates of $a(t)$ using these scaled parameters (Extended Data Fig. 3a, b). We also tested whether different neural response lags would affect the results. To do this, we compared neural tuning curves for PPC and FOF while varying the latency between auditory clicks and neural representation for each area using a range from 0.1 to 0.25 s. This results in a matrix of comparisons, where response lag for PPC varies along one axis and response lag for FOF varies along the other axis (Extended Data Fig. 3d).

Optical fibre chemical sharpening. Modifying previously published protocols[21], we began construction with a standard off the shelf 50/125 µm FC-FC duplex fibre cable (http://www.fibercables.com). The cable jacket, strengthening fibres, and outer plastic coating (typically white or orange) were fully removed leaving 1 cm of fibre optic cable and inner plastic coating (typically clear) intact. Then 2 mm of the fibre tip (with final layer of plastic coating still attached) was submerged in 48% hydrofluoric acid topped with mineral oil for 85 min, followed by water for 5 min (submerging 5 mm), and acetone for 2 min (to soften the plastic). The plastic coating was then gently cut with a razor and pulled off with tweezers to reveal a 1 mm sharp-etched fibre tip. Enough plastic was removed, depending on the depth of the targeted site, to ensure that only the glass fibre optic would be inserted into the brain.

Optogenetic virus injection and fibre implantation. For optogenetic perturbation experiments, the general surgery techniques follow previous reports[13]. Here we describe the techniques that were unique for this experiment. For viral injection, 2 µl of adeno-associated virus (AAV) (AAV5-CaMKIIα-eNpHR3.0-eYFP[22]) was lightly dyed with fast green powder and front loaded into a glass pipette mounted to a Nanoject (Drummond Scientific) prefilled with mineral oil. The pipette tip was manually cut to ~30 µm diameter. Five closely spaced injection tracts were used with each animal. For the central injection tract, two injections of 9.2 nl were made every 100 µm in depth starting 200 µm below brain surface for FOF for 1.5 mm. Four additional injection tracts were completed, using procedures identical to the central tract, one 500 µm anterior, posterior, medial and lateral from the central tract. Each injection was followed by a 10 s pause, with 1 min following the final injection in a tract before the pipette was removed. A total of 1.5 µl of virus was injected over a 30-min period consisting of ~160 separate injections. A chemically sharpened fibre optic (50 µm core, 125 µm cladding) was then lowered down the central injection tract to a depth of 1 mm. The craniotomy was filled with kwik-sil (World Precision Instruments), allowed to set for 10 min, and the fibre optic was secured to the skull with C&B Metabond and dental acrylic. After surgery, the primary incision was closed with sutures that allowed only the FC connector to protrude. Halorhodopsin expression was allowed to develop for 6 weeks before behavioural testing began.

Optogenetic pertubation. The animal's implant was connected to a 1-m patch cable attached to a single fibre rotary joint (Princetel) mounted on the ceiling of the behavioural chamber. This was connected to a 200 mW, 532 nm laser (OEM Laser Systems) operating at 25 mW, which was triggered with a 5 V transistor–transistor logic (TTL) pulse. Laser illumination occurred on a random 25% of trials. Behavioural bias was measured as the mean difference between the observed 'go ipsi' rate for inactivation and control trials for each of 10 binned stimulus strengths. Thus, a positive value represents an increase in ipsilateral responses on laser illumination trials. Confidence intervals and statistical comparisons for this bias parameter were calculated using nonparametric bootstrap procedures resampling from the population of rats with replacement. All rats were included in the analysis of full-trial inactivation. However, two rats were not included in the higher temporal resolution optogenetic experiments because they did not exhibit significant full-trial effects. The 500-ms and 250-ms experiments were performed sequentially in different groups of rats. Therefore, the assignment of group was not randomized.

Measurement of optogenetic inactivation. To measure the effects of optogenetic inactivation, both acute and chronic neural recordings were performed during optogenetic inactivation. For the acute recordings, a sharp etched fibre optic and sharp tungsten electrode (0.5 or 1.0 MΩ) were independently advanced into the cortex of an anaesthetized rat. The fibre was positioned in the centre of the field of infection at a given depth. Every location that yielded regular spiking activity of ~5 Hz or greater (single and multi-unit activity) was tested 10 times with 500 ms periods of laser illumination at 25 mW every 5 s. For chronic recordings, a silicon probe array (A4x8-5mm-200-400-177, NeuroNexus) was lowered 1.8 mm into cortex. The probes were oriented in the medial–lateral dimension, centred at 1.3 mm medial and 2.25 mm anterior from bregma (0.25 mm anterior from the central injection tract). The optical fibre tip was lowered in at a 30° angle with the final tip position located 1 mm from the brain surface, 2 mm anterior and 1.3 mm medial from bregma. Laser illumination was identical to that used for other awake, behaving rats.

Analysis of optogenetic effects using behavioural model. We used the trial-by-trial behavioural model described previously (Extended Data Fig. 2) to understand better the cause of the ipsilateral bias induced by unilateral FOF inactivation (Fig. 4a–c). The original model had only one parameter to describe a right versus left choice bias, the decision borderline þ. Following J.C.E. (manuscript in preparation) we adapted the model, adding several parameters that could describe possible sources of a side bias, and asked which of these variants best fit the rats' behaviour during trials in which the FOF was unilaterally inactivated. The four different sources of a choice bias that we considered were:

(1) Post-categorization bias. The model's distribution of values of the accumulator a at the end of the stimulus is categorized into 'go right' versus 'go left' choices according to $a >$ þ versus $a <$ þ, in which þ is the decision borderline parameter. When performing unilateral inactivation, the choice directions can be mapped as 'contralateral' or 'ipsilateral' with respect to the side of inactivation. We defined post-categorization bias as a process that takes a fraction of the trials categorized as choices contralateral to the inactivated side of the brain, and converts them into ipsilateral choices. The fraction of trials thus converted is a free parameter, fit to the behavioural data (from unilateral FOF inactivation trials).

(2) Accumulator shift. This is an additive bias that alters the final value of the accumulator. It is mathematically equivalent to altering the value of the decision borderline þ that categorizes accumulator values into right and left choices (or, contralateral and ipsilateral when considering the choice with respect to the side of unilateral inactivation). In the behavioural model, this is implemented by changing a to a + shift after the end of accumulation but before the application of the decision borderline, or (equivalently) by changing the decision borderline þ to þ − shift, with shift being the free parameter, fit to the behavioural data from unilateral FOF inactivation trials.

Because the FOF has been suggested as analogous or even homologous to the primate FEF[13,23–25], and the primate FEF is known to be involved in spatial attention[26], we considered two possible ways in which an effect that could be interpreted as unbalanced spatial attention might bias sensory input signals during the task:

(3) Unbalanced input gain. In the unperturbed case, unadapted right and left clicks both have the same effect magnitude on the accumulator a (+1 and −1 for right and left clicks, respectively, before accounting for click adaptation). We considered the possibility that the clicks from the two sides could have different impact magnitudes. Such an unbalanced input gain was implemented by allowing the contralateral click magnitude to be a free parameter (while preserving its sign), and this free parameter was fit to the behavioural data from unilateral FOF inactivation trials.

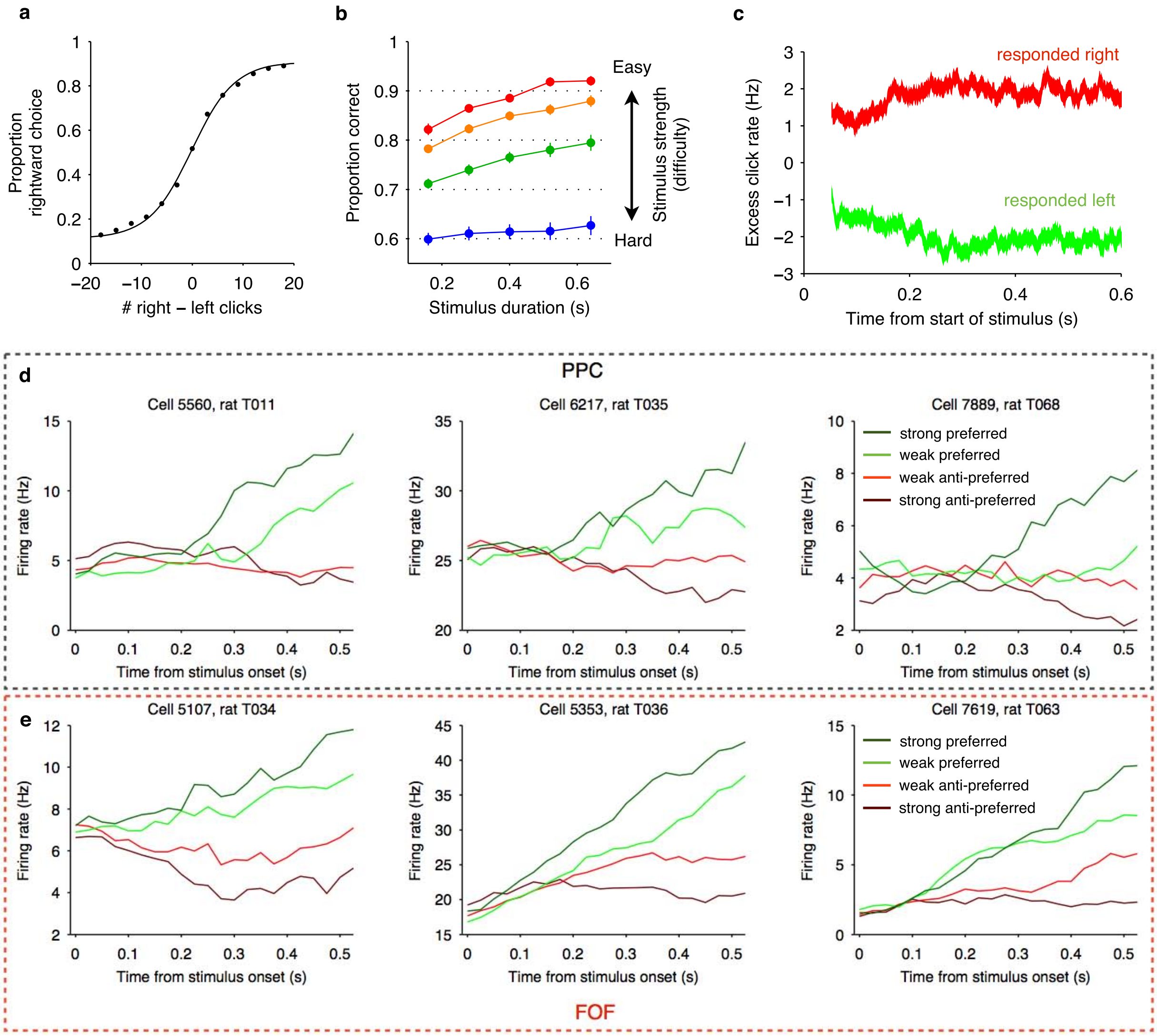
(4) Unbalanced input noise. In the unperturbed case unadapted right and left clicks both contribute the same amount of variance to the accumulator a. Thus both right and left click streams have the same signal-to-noise ratio (SNR). Here we considered varying the SNR of one side by altering the variance contributed to the accumulation by ipsilateral versus contralateral clicks. The unbalanced input noise was implemented by allowing the contralateral noise variance to be a free parameter, fit to the behavioural data from unilateral FOF inactivation trials.

We note that (3) and (4), unbalanced input gain and unbalanced input noise, both indicate mechanisms that act throughout the entirety of the accumulation process. By contrast, (1) and (2), the post-categorization bias and the accumulator shift, can be implemented via a mechanism that acts only at the end of the accumulation process.

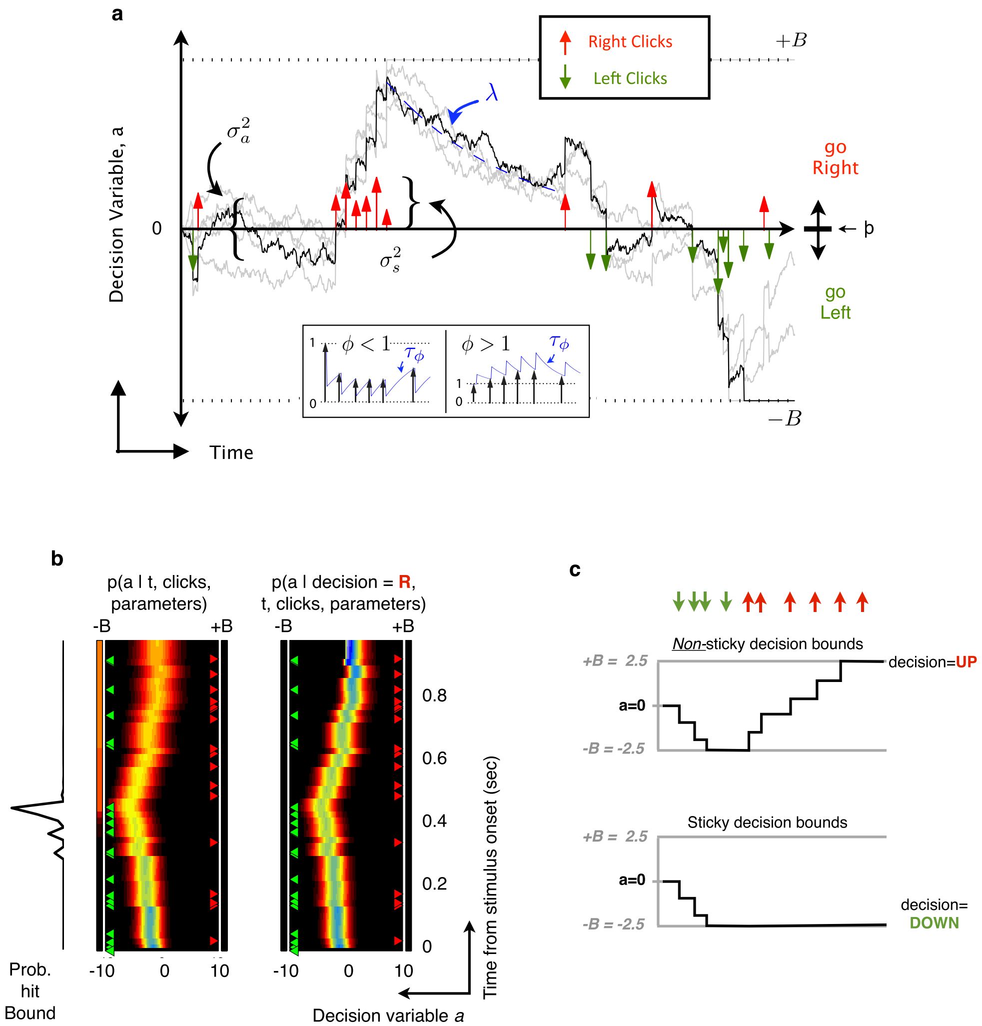
We compared how well these possible sources of bias in the behavioural model could account for the trial-by-trial choice behaviour during both full-trial and peri-choice FOF inactivation. For the full-trial inactivation, we considered all four alternative implementations described above. For the peri-choice inactivation, we considered the two alternatives that predict an effect on behaviour only during the end of the stimulus presentation period, namely (1) and (2), the post-categorization bias and

the accumulator shift. In both the full-trial and peri-choice inactivation cases, we began by fitting the control non-inactivation trials with the standard model described in Extended Data Fig. 2, combining data across all rats used in each type of experiment. These control non-inactivation trials comprised 75% of the total trials and were randomly interleaved with the inactivation trials within the same behavioural sessions. Starting from the parameters that best fit behaviour on these control trials, we then examined how well each single-parameter implementation of bias under consideration could fit the inactivation data. In all cases, we report both the overall log-likelihood of the fit, and how well, on average, each alternative source of bias predicts the response on each trial. We illustrate the fits across multiple trials by comparing the psychometric functions produced by the different models to those found in the data. We performed statistical comparisons of the quality of the fits using a nonparametric resampling bootstrap procedure. We first generated 200 sets of trials with the same trial count as the original data by resampling the original data with replacement. We then fit each of these sets with the models described above and measured the difference in likelihood between post-categorization bias and the other alternatives for each resampled set of data. This provided a distribution of likelihood ratios from which we could estimate the statistical probability that post-categorization bias provides an improved fit compared to the other models.

20. Guo, Z. V. *et al.* Flow of cortical activity underlying a tactile decision in mice. *Neuron* **81,** 179–194 (2014).
21. Lambelet, P., Sayah, A., Pfeffer, M., Philipona, C. & Marquis-Weible, F. Chemically etched fiber tips for near-field optical microscopy: a process for smoother tips. *Appl. Opt.* **37,** 7289–7292 (1998).
22. Gradinaru, V. *et al.* Molecular and cellular approaches for diversifying and extending optogenetics. *Cell* **141,** 154–165 (2010).
23. Leonard, C. M. The prefrontal cortex of the rat. I. Cortical projection of the mediodorsal nucleus. II. Efferent connections. *Brain Res.* **12,** 321–343 (1969).
24. Neafsey, E. J. *et al.* The organization of the rat motor cortex: a microstimulation mapping study. *Brain Res.* **11,** 77–96 (1986).
25. Guandalini, P. The corticocortical projections of the physiologically defined eye field in the rat medial frontal cortex. *Brain Res. Bull.* **47,** 377–385 (1998).
26. Squire, R. F., Noudoost, B., Schafer, R. J. & Moore, T. Prefrontal contributions to visual selective attention. *Annu. Rev. Neurosci.* **36,** 451–466 (2013).



Extended Data Figure 1 | Rat behaviour and example neurons. **a**, Mean psychometric function across all rats. Accuracy was highest for the largest click differences (the left and right endpoints of the curve) and lower for smaller click differences (the middle of the curve). **b**, Mean chronometric function across all rats. Trials were sorted by binned stimulus strength (difficulty), with mean click ratios ranging from 39:1 clicks s^{-1} for the easiest trials to 25:15 clicks s^{-1} for the hardest trials. In general, accuracy improved with longer stimulus durations. **c**, Mean psychophysical reverse correlation across all rats. This was calculated based on trials with minimum duration of at least 0.6 s. For each time point in each trial, we first computed the excess click rate difference (right − left clicks s^{-1}) relative to the value expected given the random processes used to generate the trial. These excess click rates were averaged separately for trials ending with a right choice (red) and for trials ending with a left choice (green). The separation between the two traces indicates how strongly clicks from the corresponding time point influence the final decision. **d**, Peri-event time histograms (PETHs) aligned to stimulus onset were calculated for three example PPC neurons. Trials were sorted into four stimulus strength bins for each neuron. Green traces correspond to the preferred-direction stimuli and red traces correspond to anti-preferred-direction stimuli. Darker colours correspond to stronger stimuli (less difficult) and brighter colours correspond to weaker stimuli (more difficult). **e**, PETHs for three example FOF neurons using the same conventions. In both regions, individual neurons exhibit ramping activity that depends on stimulus strength.



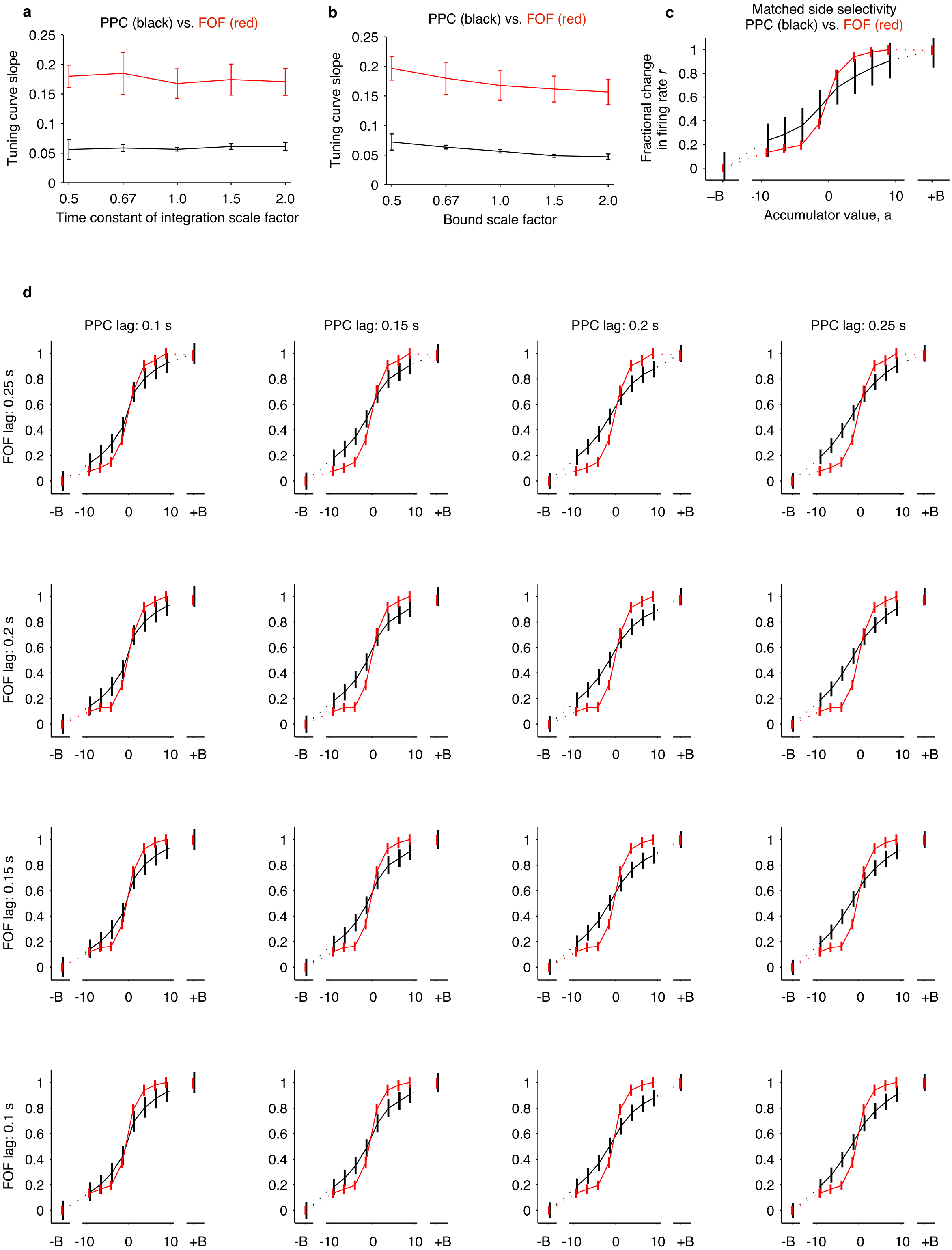
Extended Data Figure 2 | Behavioural model (ref. 10). a, At each time point, the accumulator a (black trace) represents an estimate of the 'right' versus 'left' evidence accrued so far. At stimulus end, the model decides 'right' if $a >$ þ (the decision borderline) and 'left' otherwise, in which þ is a free parameter. Light grey traces indicate alternative runs with different instantiations of model noise. These example trajectories are for illustrative purposes; the model estimates the full probability distribution of a at each time point. Right ↑ (left ↓) clicks change the value of a by positive (negative) impulses of magnitude C. σ_a^2 is a diffusion constant, parameterizing noise in a. σ_s^2 parameterizes noise associated with each click. λ parameterizes consistent drift in the decision variable a. In the 'leaky' case ($\lambda < 0$, illustrated), drift is towards $a = 0$, and later clicks impact the decision more than earlier clicks. In the 'unstable' case ($\lambda > 0$), drift is away from $a = 0$, and earlier clicks effect the decision more than later clicks. The time constant of the accumulation process is $\tau = 1/\lambda$. B is the height of sticky decision bounds. If a reaches either bound, it leads to decision commitment before the end of the stimulus and later clicks have no effect on the choice. ϕ, τ_ϕ parameterize sensory adaptation by defining the dynamics of C. Immediately after a click, the magnitude C is multiplied by ϕ. C then recovers towards an unadapted value of 1 with time constant τ_ϕ. Facilitation thus corresponds to $\phi > 1$, while depression corresponds to $\phi < 1$. These properties are implemented by the following equations: if $|a| \geq B$ then $da/dt = 0$; else

$$da = \sigma_a dW + (\delta_{t,t_R} \cdot \eta_R \cdot C - \delta_{t,t_L} \cdot \eta_L \cdot C)dt + \lambda a dt$$

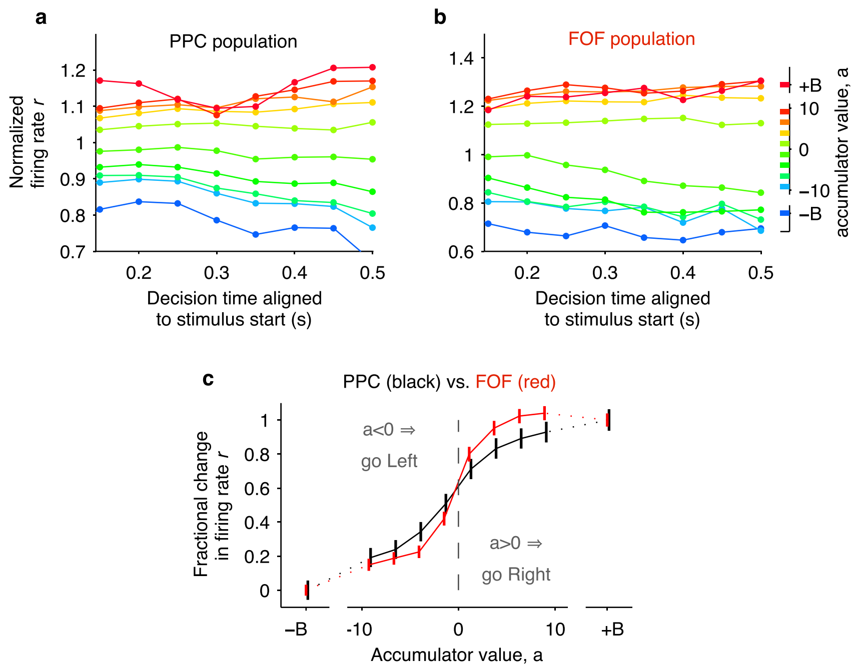
$\delta_{t,t_{R,L}}$ are delta functions at the times of the auditory clicks. η are independent and identically distributed (i.i.d.) Gaussian variables drawn from $N(1,\sigma_s)$. dW is a white-noise Wiener process. Adaptation dynamics are given by:

$$\frac{dC}{dt} = \frac{1-C}{\tau_\phi} + (\phi - 1)C(\delta_{t,t_R} + \delta_{t,t_L})$$

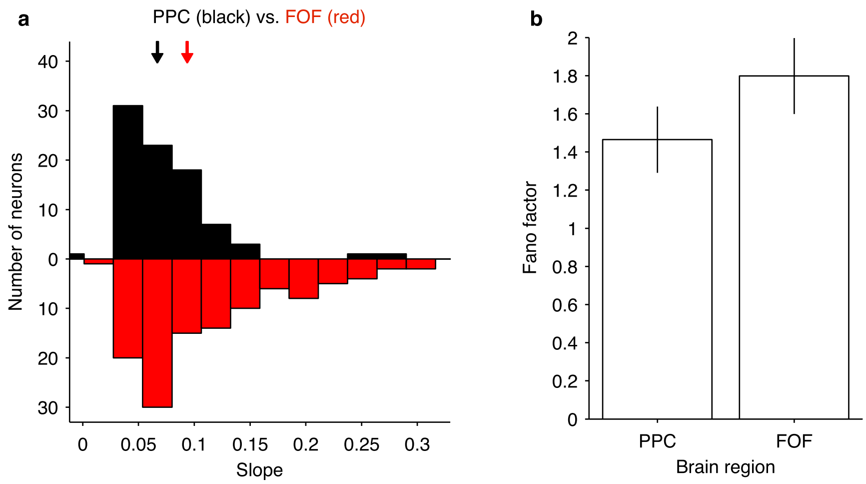
In addition, a lapse rate parameterizes the fraction of trials on which a random response is made. **b**, The behavioural model provides an estimate of the evolution of the distribution of a for each trial, with colour representing probability density for both panels. The forward version of the model (left panel) estimates the distribution of a at each time point based entirely on the click times and model parameters. Leftward (green) clicks push the distribution more negative and rightward (red) clicks push it more positive. The final value obtained by a at the end of the trial dictates the choice. In this example, the distribution of the final value of a is more heavily weighted towards negative values because there were more leftward than rightward clicks for this trial. A better estimate of the distribution of a can be obtained by also taking into account the subject's choice made at the end of the trial (right panel). The final choice constrains the distribution of a at the final time point to have all its mass to one side of the decision boundary (in this example trial, despite the many leftward clicks, the subject chose right, and thus at the final time point, all the probability mass is at $a > 0$). This constraint is then propagated backwards in time, to obtain the distribution of accumulator values at each time point that is consistent both with the stimulus clicks and the subject's final choice. The final result is an estimate of the distribution of a at each time point that takes into account the click times, the model parameters, and the subject's choice. **c**, Illustration of non-sticky versus sticky decision bounds. Top: response of an accumulator to a sequence of downward (green arrows) and upward (red arrows) impulses when the bound parameter B is 2.5 and the bounds are not sticky. The fourth downward arrow (green) has no effect, because the bounds have been reached, and a cannot go beyond them. But subsequent upward arrows (red) do have an effect, because they do not push against the bounds. Bottom: response of an accumulator with the same parameters, receiving the same sequence of impulses, but when the bounds are sticky. Impulses subsequent to reaching the bound have no effect. We fit our rats' behaviour data to this non-sticky bound model. We found that, similar to the version with the sticky bounds, the accumulation time constants were long ($|\tau| = 1/|\lambda| = 1.0 \pm 0.2$ s, mean ± s.e.m. across rats), and the bounds were high (17.1 ± 2.2, mean ± s.e.m. across rats). Such high bounds once again indicate that the bounds have minimal impact (consistent with this, the difference between the sticky and non-sticky bound models turned out to be negligible).



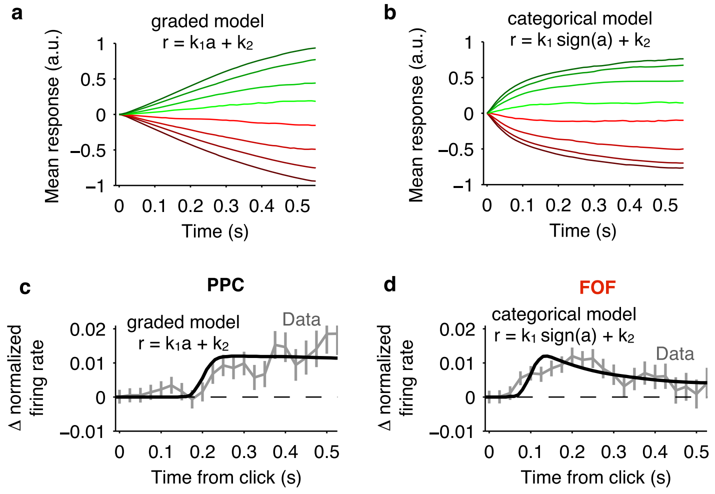
Extended Data Figure 3 | Robustness of distinction between PPC and FOF to parameter variation. **a**, The distinction between PPC and FOF encoding is robust to variation of the model's time constant of integration. The slope of the tuning curves, drawn from the same analysis as Fig. 3c in the main text, except that here the analysis was carried out at a variety of integration time constants. **b**, Same analysis performed for a variety of heights for the sticky decision-commitment bounds. In both cases, the corresponding best-fit parameter was scaled by the factor shown on the horizontal axis, and the slope of tuning curve for FOF (in red) versus PPC (in black) was plotted as a function of that scale factor. Error bars show 95% confidence intervals. The slope of the FOF tuning curve is significantly sharper than the slope of the PPC tuning curve across the entire range of parameter values tested ($P < 0.05$). **c**, Tuning curve comparison between PPC and FOF with subset of PPC neurons selected such that the two regions have matched side selectivity. This resulted in $n = 50$ neurons for PPC and the original $n = 128$ neurons for FOF. The tuning curve is significantly steeper for FOF ($P < 0.05$). **d**, The same analysis as Fig. 3c in the main text, except that here we varied the latency applied between click time and neural representation (see Methods). While we would expect that an improper choice of latency would degrade the quality of the estimate of the accumulator value, the slope at the zero-crossing was still significantly larger for FOF compared to PPC for all comparisons ($P < 0.05$).



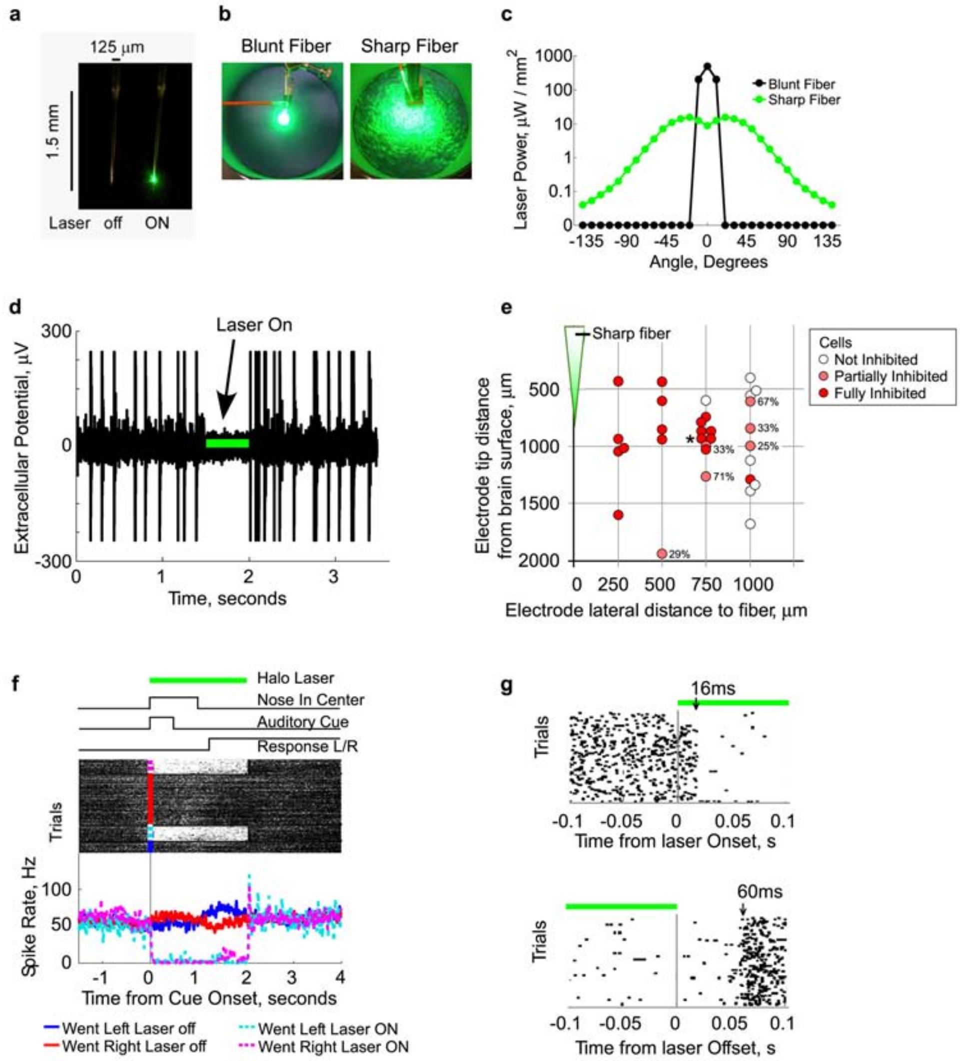
Extended Data Figure 4 | Tuning curves based on model fit to aggregate rat behaviour. The same analysis as Fig. 3 in the main text, except that rather than fitting the behavioural model on a rat-by-rat basis, it was fit to the aggregate behaviour of all rats. Thus, all rats share identical model parameter values for this figure. Although we would expect that to degrade the quality of the estimate of the accumulator value, the slope at the zero-crossing was still significantly larger for FOF than for PPC (PPC slope = 0.079 ± 0.004, FOF slope = 0.135 ± 0.026, mean ± 95% confidence interval).



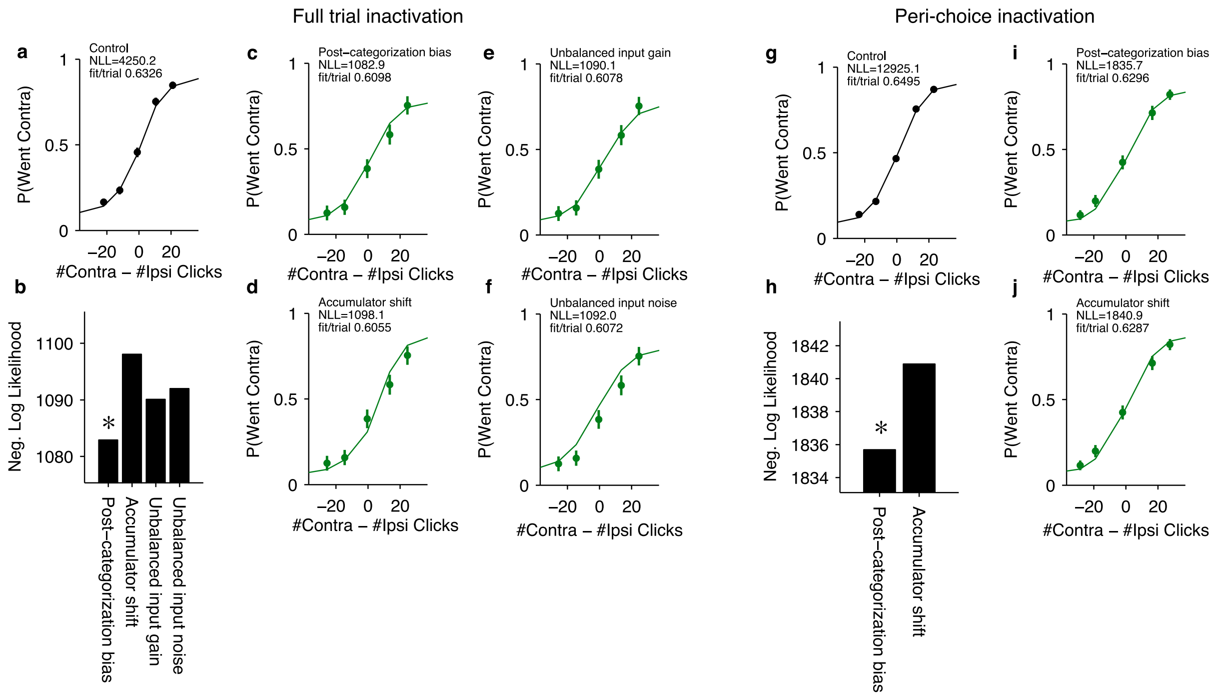
Extended Data Figure 5 | Individual neuron analyses. **a**, The mapping between firing rate and accumulator value is shifted towards steeper slopes for the distribution of FOF neurons compared to the distribution of PPC neurons. Bars show the histogram of individual neuron slopes obtained from a sigmoidal fit of the relationship between firing rate and accumulator value, averaging across time from 0.15 to 0.5 s. Black bars correspond to PPC and red bars correspond to FOF. The arrows indicate the medians of the two distributions. While there is considerable overlap between the two populations, there is a significant shift towards greater steepness for the distribution of FOF neurons compared to PPC neurons ($P < 0.001$, rank sum test). A larger slope corresponds to a steeper change from low to high firing rates at the transition between negative and positive accumulator values. Thus, a steeper slope is associated with a more categorical as opposed to graded encoding of the accumulating evidence. **b**, To compare neural variability, we measured the fano factor for each neuron as a function of the accumulated evidence and compared across regions. There was not a significant difference in neural variability for the representations in the two areas ($P = 0.23$).



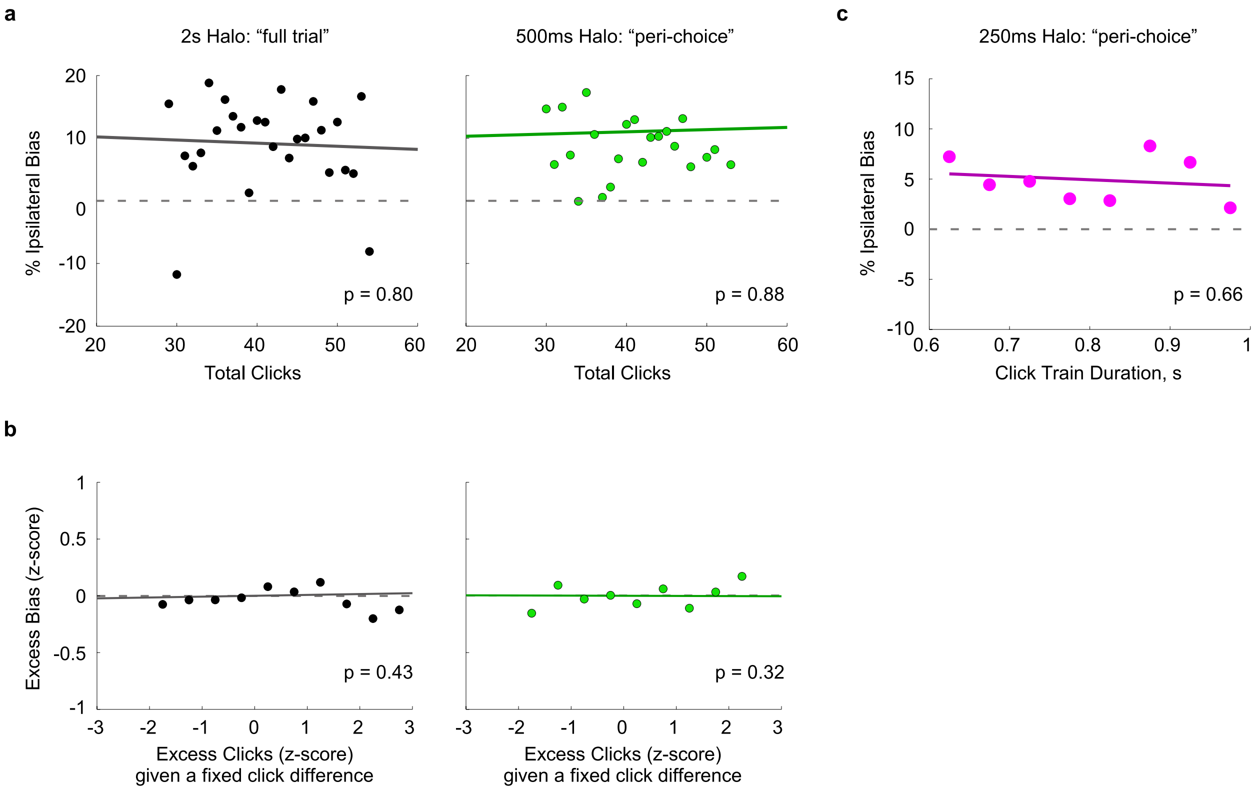
Extended Data Figure 6 | Expected response profiles and click-triggered averages based on graded and categorical encodings. **a**, Using stimuli corresponding to what was presented to the rats, individual trial responses $r(t)$ were calculated based on the graded encoding $r(t) = r = k_1a(t) + k_2$, in which $a(t)$ is the decision variable obtained from the behavioural model. Then, responses were averaged across trials sorting based on mean stimulus strength with the same exact method as described for Fig. 1. **b**, Same as **a** for a categorical encoding $r(t) = k_1\text{sign}(a(t))+k_2$. The ratio of the positive-to-negative encoding changes over time for each condition, leading to a ramping response profile. This highlights the point that ramping response profiles alone are not a sufficient demonstration of a gradual accumulation process. **c**, The black line shows the model prediction for the click-triggered average based on a graded encoding of accumulated evidence ($r = k_1a + k_2$) with a 0.2 s mean encoding lag to match the response lag of PPC. The grey line shows the data from Fig. 1. **d**, Same as **c** for a categorical encoding of accumulated evidence ($r = k_1\text{sign}(a) + k_2$) with a 0.1 s mean encoding lag to match the response lag of FOF. For both schemes, the encoding lag was taken from a Gaussian distribution with a 0.02 s standard deviation to account for variability in lag across neurons.



Extended Data Figure 7 | Chemically sharpened fibre optics allow extensive inhibition during acute and chronic recordings from cortical regions expressing eNpHR3.0. **a**, Image of a chemically sharpened 50 µm core, 125 µm cladding fibre. **b**, Light spot produced by a blunt and sharpened fibre 2 cm above the floor of a cylindrical container 10.5 cm in diameter. **c**, Laser power output from a blunt and sharpened fibre as a function of angle relative to the fibre optic tip. 25 mW input power. Power meter was 2.86 mm from the fibre tip. **d**, Single trace of an acute recording of spontaneous activity in anesthetized primary somatosensory cortex (S1, 1.5 mm posterior, 2.8 mm lateral from Bregma) expressing eNpHR3.0. Laser illumination period, 500 ms, marked by the green bar. **e**, Location of acute recording units (single and multi) in anaesthetized S1 relative to fibre tip and cortical surface. The level of inhibition was measured from ten repeated 500 ms laser illumination periods, delivered every 5 s. Percentage reduction displayed next to partially inhibited units. Unit in **d** indicated with an asterisk. **f**, Example multiunit activity from the FOF of a rat performing a memory guided orienting task. The 2-s laser illumination period initiated at cue onset, resulted in 97% inhibition of spiking activity for both trials where the rat made left or made right responses. **g**, Multiunit spiking activity (from **f**) aligned to laser onset (top) and laser offset (bottom). Spiking activity is strongly inhibited ~16 ms after laser onset and recovers ~60 ms after laser offset.



Extended Data Figure 8 | Model-based analysis of halorhodopsin-mediated inactivation of the FOF (following J.C.E. *et al.*, manuscript in preparation). **a**–**f**, Full-trial inactivation (see Fig. 4a). **a**, Control trials with no inactivation taken from sessions with full-trial, 2 s, halorhodopsin inactivation. 'Contra' and 'ipsi' sides are relative to the side of the FOF that was inactivated on non-control trials. The curve shows the psychometric function predicted by the best-fit behavioural model based on the same stimuli that produced the behaviour. **b**, Comparison of the negative log likelihood for each candidate source of bias (see Methods). Smaller values correspond to a better fit. The post-categorization bias was significantly better than all other alternatives ($P < 0.05$, bootstrap). **c**–**f**, Data points show the proportion of contralateral choices for full-trial, 2 s inactivation of FOF. The curves show choice behaviour predicted by each alternative implementation of choice bias. **b**, Post-categorization bias. **c**, Accumulator shift. **d**, Unbalanced input gain. **e**, Unbalanced input noise. **g**–**j**, Peri-choice inactivation (Fig. 4b–d of main text). **g**, Control trials with no inactivation taken from sessions with either 500 ms or 250 ms peri-choice inactivation. The curve shows the psychometric function predicted by the best-fit behavioural model based on the same stimuli that produced the behaviour. **h**, Comparison of the negative log likelihood for both sources of bias. Smaller values correspond to a better fit. The post-categorization bias was significantly better than the accumulator bias ($P < 0.05$, bootstrap). **i**, **j**, Data points show the proportion of contralateral choices for peri-choice inactivation of FOF. The curves show choice behaviour predicted by the two versions of bias that predict an effect only at the end of the stimulus period. **i**, Post-categorization bias. **j**, Accumulator shift.



Extended Data Figure 9 | FOF inactivation induced response bias does not correlate with click count or click train duration. This figure shows further analysis of the data from Fig. 4 of the main text. **a**, **b**, During both the full-trial and 500 ms peri-choice inactivation sessions 50% of the trials had the click train duration fixed at 1 s (half inactivation half control trials). The remaining 50% were non-inactivation control trials with stimulus durations that varied randomly (not included in this analysis). All trial types were randomly interleaved. The Poisson nature of the click stimuli means that the number of clicks varied from trial to trial even for trials with the same duration and net click difference. **a**, We first asked whether the magnitude of the ipsilateral bias was correlated with the number of clicks for these fixed duration trials, and found that it was not ($P = 0.80$ and $P = 0.88$, for full trial and 500 ms peri-choice inactivations, respectively). **b**, In a separate analysis we asked whether given a fixed click difference, do more clicks lead to an excess bias. We first separated trials into groups with equivalent click differences (ipsi count − contra count). The actual click count of each trial was then subtracted by its group mean and normalized by its group standard deviation. This gives the *z*-score of the excess clicks on each trial given the click difference of that trial. We repeated this *z*-score normalization for the response bias on each trial (1 for respond ipsi, 0 for respond contra, bias = response on an inactivation trial − mean response on equivalent non-inactivation control trials). This gives a plot of the excess bias as a function of excess clicks. We found no significant correlation ($P = 0.43$ and $P = 0.32$ for full trial and 500 ms peri-choice inactivations, respectively). **c**, In the 250-ms inactivation experiment, click train durations varied both for control and inactivation trials. Here we asked whether the induced bias was correlated with the variable click train duration, and again found that it was not ($P = 0.66$).

Extended Data Table 1 | Best fit parameters of behavioural model

Experiment	Data set	λ	σ_a^2	σ_s^2	B	ϕ	τ_ϕ	þ	lapse	fit / trial
PPC recordings	B053	1.51	0.0	23.3	15.0	0.39	0.07	0.02	0.07	0.47
PPC recordings	T011	1.44	0.0	61.8	19.9	0.45	0.02	-0.58	0.05	0.40
PPC recordings	T035	-1.61	24.6	178.7	26.2	0.32	0.03	-0.15	0.05	0.58
PPC/FOF recordings	T068	2.45	0.0	57.5	11.6	0.36	0.04	0.99	0.11	0.46
FOF recordings	B068	2.17	0.0	16.0	16.8	0.41	0.17	0.14	0.10	0.57
FOF recordings	T030	0.46	10.1	39.0	14.5	0.32	0.03	0.20	0.04	0.38
FOF recordings	T034	-0.99	49.7	10.6	17.2	0.58	0.07	-0.69	0.20	0.55
FOF recordings	T036	0.84	0.0	34.1	14.6	0.51	0.07	-0.62	0.11	0.48
FOF recordings	T063	0.42	2.5	139.3	19.6	0.43	0.03	0.09	0.01	0.51
PPC/FOF recordings	physiology combined rat	0.47	0.0	61.3	19.2	0.56	0.06	-0.13	0.09	0.49
FOF optogenetics	B108	0.44	0.1	52.3	17.3	0.10	0.04	-0.57	0.19	0.50
FOF optogenetics	B116	-0.14	54.6	42.5	30.5	0.39	0.05	0.01	0.04	0.47
FOF optogenetics	B117	-2.08	66.0	199.8	31.6	0.48	0.01	0.06	0.00	0.50
FOF optogenetics	B119	-1.65	116.3	88.9	10.2	0.36	0.02	-0.57	0.11	0.54
FOF optogenetics	B120	0.62	0.0	138.4	31.8	0.94	0.50	0.10	0.00	0.42
FOF optogenetics	B123	-1.88	7.0	99.3	31.7	0.54	0.04	-0.58	0.07	0.46
FOF optogenetics	B124	-1.11	0.0	131.3	28.0	0.61	0.03	0.87	0.05	0.44
FOF optogenetics	B125	-0.71	74.4	90.9	11.2	0.30	0.02	-0.02	0.02	0.47
FOF optogenetics	B126	-0.34	0.0	111.0	27.0	0.93	0.50	-0.11	0.00	0.40
FOF optogenetics	B129	0.79	16.5	6.2	15.6	0.29	0.06	0.11	0.05	0.37
FOF optogenetics	K145	0.64	27.6	49.3	13.5	0.45	0.06	-0.26	0.18	0.51
FOF optogenetics	K147	0.07	25.8	3.9	17.8	0.57	0.16	-0.63	0.21	0.52
FOF optogenetics	K149	0.80	46.0	199.6	31.6	0.26	0.01	1.39	0.11	0.50
FOF optogenetics	K150	0.25	25.0	11.9	18.0	0.34	0.09	0.19	0.16	0.51
FOF optogenetics	K151	0.35	0.0	78.8	18.2	0.39	0.05	-0.11	0.07	0.48
FOF optogenetics	K156	-2.39	0.7	157.0	31.1	0.43	0.02	0.64	0.05	0.48
FOF optogenetics	K157	-0.30	32.7	37.4	23.0	0.29	0.05	-0.68	0.11	0.48
FOF optogenetics	K158	-0.58	39.6	61.1	24.3	0.92	0.50	0.19	0.12	0.46
FOF optogenetics	full trial inactivation controls, combined rat	0.10	0.0	132.6	16.0	0.04	0.06	0.00	0.07	0.63
FOF optogenetics	peri-choice inactivation controls, combined rat	-0.39	0.0	97.7	15.2	0.21	0.11	0.01	0.06	0.65

doi:10.1038/nature14101

Common genetic variants influence human subcortical brain structures

A list of authors and their affiliations appears at the end of the paper

The highly complex structure of the human brain is strongly shaped by genetic influences[1]. Subcortical brain regions form circuits with cortical areas to coordinate movement[2], learning, memory[3] and motivation[4], and altered circuits can lead to abnormal behaviour and disease[2]. To investigate how common genetic variants affect the structure of these brain regions, here we conduct genome-wide association studies of the volumes of seven subcortical regions and the intracranial volume derived from magnetic resonance images of 30,717 individuals from 50 cohorts. We identify five novel genetic variants influencing the volumes of the putamen and caudate nucleus. We also find stronger evidence for three loci with previously established influences on hippocampal volume[5] and intracranial volume[6]. These variants show specific volumetric effects on brain structures rather than global effects across structures. The strongest effects were found for the putamen, where a novel intergenic locus with replicable influence on volume (rs945270; $P = 1.08 \times 10^{-33}$; 0.52% variance explained) showed evidence of altering the expression of the *KTN1* gene in both brain and blood tissue. Variants influencing putamen volume clustered near developmental genes that regulate apoptosis, axon guidance and vesicle transport. Identification of these genetic variants provides insight into the causes of variability in human brain development, and may help to determine mechanisms of neuropsychiatric dysfunction.

At the individual level, genetic variations exert lasting influences on brain structures and functions associated with behaviour and predisposition to disease. Within the context of the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium, we conducted a collaborative large-scale genetic analysis of magnetic resonance imaging (MRI) scans to identify genetic variants that influence brain structure. Here, we focus on volumetric measures derived from a measure of head size (intracranial volume, ICV) and seven subcortical brain structures corrected for the ICV (nucleus accumbens, caudate, putamen, pallidum, amygdala, hippocampus and thalamus). To ensure data homogeneity within the ENIGMA consortium, we designed and implemented standardized protocols for image analysis, quality assessment, genetic imputation (to 1000 Genomes references, version 3) and association (Extended Data Fig. 1 and Methods).

After establishing that the volumes extracted using our protocols were substantially heritable in a large sample of twins ($P < 1 \times 10^{-4}$; see Methods and Extended Data Fig. 11a), with similar distributions to previous studies[1], we sought to identify common genetic variants contributing to volume differences by meta-analysing site-level genome-wide association study (GWAS) data in a discovery sample of 13,171 subjects of European ancestry (Extended Data Fig. 2). Population stratification was controlled for by including, as covariates, four population components derived from standardized multidimensional scaling analyses of genome-wide genotype data conducted at each site (see Methods). Site-level GWAS results and distributions were visually inspected to check for statistical inflation and patterns indicating technical artefacts (see Methods).

Meta-analysis of the discovery sample identified six genome-wide significant loci after correcting for the number of variants and traits analysed ($P < 7.1 \times 10^{-9}$; see Methods): one associated with the ICV, two associated with hippocampal volume, and three with putamen volume. Another four loci showed suggestive associations ($P < 1 \times 10^{-7}$) with putamen volume (one locus), amygdala volume (two loci), and caudate volume (one locus; Table 1, Fig. 1 and Supplementary Table 5). Quantile–quantile plots showed no evidence of population stratification or cryptic relatedness (Extended Data Fig. 4a). We subsequently attempted to replicate the variants with independent data from 17,546 individuals.

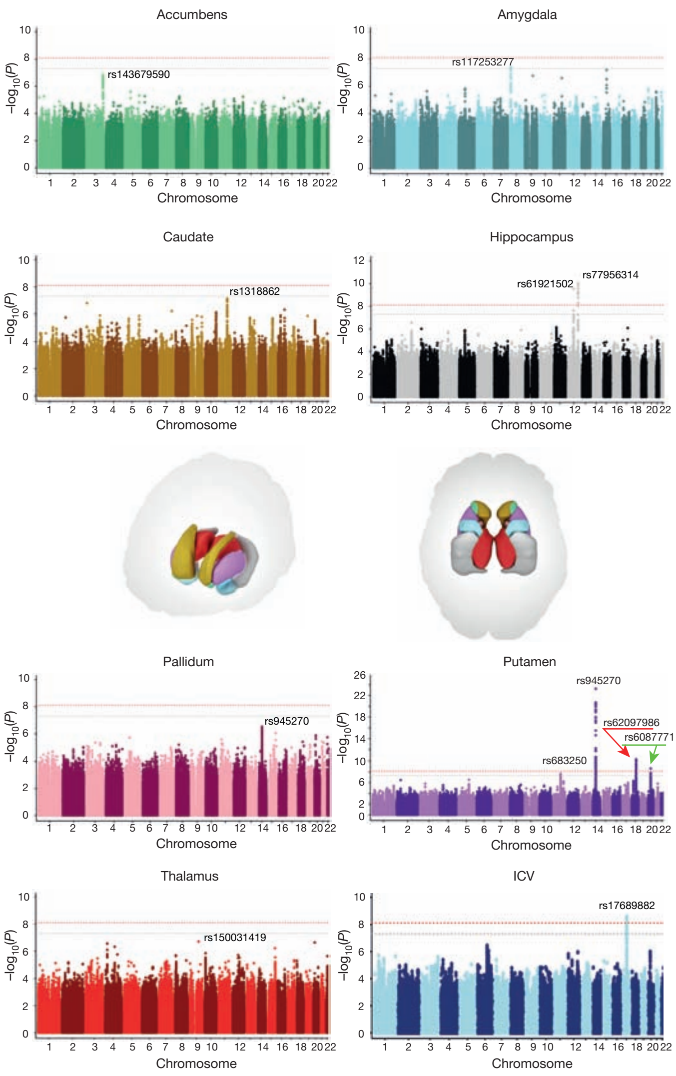


Figure 1 | Common genetic variants associated with subcortical volumes and the ICV. Manhattan plots coloured with a scheme that matches the corresponding structure (middle) are shown for each subcortical volume studied. Genome-wide significance is shown for the common threshold of $P = 5 \times 10^{-8}$ (grey dotted line) and also for the multiple comparisons-corrected threshold of $P = 7.1 \times 10^{-9}$ (red dotted line). The most significant SNP within an associated locus is labelled.

Table 1 | Genetic variants at eight loci were significantly associated with putamen, hippocampus, caudate nucleus and ICV

Trait	Marker	A1	A2	Frq	Discovery cohort Effect (se)	Discovery cohort *P* value	Discovery cohort Sample size	Replication cohort Effect (se)	Replication cohort *P* value	Replication cohort Sample size	Discovery + replication cohorts Effect (se)	Discovery + replication cohorts *P* value	Total sample size	Variance explained (%)	Diff./allele (%)
Putamen	rs945270	C	G	0.58	60.64 (6.00)	5.43×10^{-24}	13,145	39.15 (5.46)	7.81×10^{-13}	15,130	48.89 (4.04)	1.08×10^{-33}	28,275	0.52	0.94
Putamen	rs62097986	A	C	0.44	39.53 (6.01)	4.86×10^{-11}	13,145	22.46 (5.53)	4.89×10^{-5}	14,891	30.28 (4.07)	1.01×10^{-13}	28,036	0.20	0.58
Putamen	rs6087771	T	C	0.71	40.72 (6.82)	2.42×10^{-9}	11,865	26.97 (6.57)	4.02×10^{-5}	13,675	33.58 (4.73)	1.28×10^{-12}	25,540	0.20	0.64
Putamen	rs683250	A	G	0.63	−33.97 (6.08)	2.33×10^{-8}	13,145	−22.30 (5.89)	1.50×10^{-4}	13,113	−27.95 (4.23)	3.94×10^{-11}	26,258	0.17	0.51
Caudate	rs1318862	T	C	0.58	26.27 (4.89)	7.54×10^{-8}	13,171	31.82 (14.23)	0.025	1,860	26.86 (4.62)	6.17×10^{-9}	15,031	0.22	0.74
Hip.	rs77956314	T	C	0.91	−54.21 (8.37)	9.33×10^{-11}	13,163	−57.43 (12.69)	6.04×10^{-6}	4,027	−55.18 (6.99)	2.82×10^{-15}	17,190	0.36	1.40
Hip.	rs61921502	T	G	0.84	43.40 (6.89)	2.92×10^{-10}	13,163	26.81 (13.32)	0.044	3,046	39.90 (6.12)	6.87×10^{-11}	16,209	0.26	1.01
ICV	rs17689882	A	G	0.22	−15,335.88 (2,582.20)	2.87×10^{-9}	10,944	−5,202.15 (5,428.60)	0.337	1,878	−13,460.47 (2,331.05)	7.72×10^{-9}	12,822	0.26	0.96

The allele frequency (frq) and effect size are given with reference to allele 1 (A1). Effect sizes are given in units of mm^3 per effect allele. Results are provided for the discovery samples and the combined meta-analysis of the discovery and replication cohorts (all European ancestry). Additional validation was attempted in non-European ancestry generalization samples (shown in Supplementary Table 6). The variance explained gives the percentage variance explained by a given SNP after correcting for covariates (see Methods for additional details). The percentage difference in volume per effect allele (Diff./allele) is based on the absolute value of the final combined effect divided by a weighted average of the brain volume of interest across all sites in the discovery sample and then multiplied by 100. Hip, hippocampus.

All subcortical genome-wide significant variants identified in the discovery sample were replicated (Table 1). The variant associated with the ICV did not replicate in a smaller independent sample, but was genome-wide significant in a previously published independent study[6], providing strong evidence for its association with the ICV. Moreover, two suggestive variants associated with putamen and caudate volumes exceeded genome-wide significance after meta-analysis across the discovery and replication data sets (Table 1). Effect sizes were similar across cohorts ($P > 0.1$, Cochran's Q test; Extended Data Fig. 4b). Effect sizes remained consistent after excluding patients diagnosed with anxiety, Alzheimer's disease, attention-deficit/hyperactivity disorder, bipolar disorder, epilepsy, major depressive disorder or schizophrenia (21% of the discovery participants). Correlation in effect size with and without patients was very high ($r > 0.99$) for loci with $P < 1 \times 10^{-5}$, indicating that these effects were unlikely to be driven by disease (Extended Data Fig. 5a). The participants' age range covered most of the lifespan (9–97 years), but only one of the eight significant loci showed an effect related to the mean age of each cohort ($P = 0.002$; rs6087771 affecting putamen volume; Extended Data Fig. 5b), suggesting that nearly all effects are stable across the lifespan. In addition, none of these loci showed evidence of sex effects (Extended Data Fig. 5c).

In our cohorts, significant loci were associated with 0.51–1.40% differences in volume per risk allele, explaining 0.17–0.52% of the phenotypic variance (Table 1); such effect sizes are similar to those of common variants influencing other complex quantitative traits such as height[7] and body mass index[8]. The full genome-wide association results explained 7–15% of phenotypic variance after controlling for the effects of covariates (Extended Data Fig. 11). Notably, the genome-wide significant variants identified here showed specific effects on single brain structures rather than pleiotropic effects across multiple structures, despite similar developmental origins as in the case of caudate and putamen (Extended Data Fig. 6a). Nevertheless, when we subjected the subcortical meta-analysis results to hierarchical clustering, genetic determinants of the subcortical structures were mostly grouped into larger circuits according to their developmental and functional subdivisions (Extended Data Fig. 6b). Genetic variants may therefore have coherent effects on functionally associated subcortical networks. Multivariate cross-structure[9] analyses confirmed the univariate results, but no additional loci reached genome-wide significance (Extended Data Fig. 6c). The clustering of results into known brain circuits in the absence of individually significant genetic variants found in the cross-structure analysis suggests variants of small effect may have similar influences across structures. Most variants previously reported to be associated with brain structure and/or function showed little evidence of large-scale volumetric effects (Supplementary Table 8). We detected an intriguing association with hippocampal volume at a single nucleotide polymorphism (SNP) with a genome-wide significant association with schizophrenia[10] (rs2909457; $P = 2.12 \times 10^{-6}$; where the A allele is associated with decreased risk for schizophrenia and decreased hippocampal volume). In general, however, we detected no genome-wide significant association with brain structure for genome-wide significant loci that contribute risk for neuropsychiatric illnesses (Supplementary Table 9).

Of the four loci influencing putamen volume, we identified an intergenic locus 50 kilobases (kb) downstream of the *KTN1* gene (rs945270; 14q22.3; $n = 28{,}275$; $P = 1.08 \times 10^{-33}$), which encodes the protein kinectin, a receptor that allows vesicle binding to kinesin and is involved in organelle transport[11]. Second, we identified an intronic locus within *DCC* (rs62097986; 18q21.2; $n = 28{,}036$; $P = 1.01 \times 10^{-13}$), which encodes a netrin receptor involved in axon guidance and migration, including in the developing striatum[12] (Extended Data Fig. 3b). Expression of *DCC* throughout the brain is highest in the first two trimesters of prenatal development[13] (Extended Data Fig. 8b), suggesting that this variant may influence brain volumes early in neurodevelopment. Third, we identified an intronic locus within *BCL2L1* (rs6087771; 20q11.21; $n = 25{,}540$; $P = 1.28 \times 10^{-12}$), which encodes an anti-apoptotic factor that inhibits programmed cell death of immature neurons throughout the brain[14] (Extended Data Fig. 3c). Consistent with this, expression of *BCL2L1* in the striatum strongly decreases at the end of neurogenesis (24–38 post-conception weeks (PCW); Extended Data Fig. 8c), a period marked by increased apoptosis in the putamen[13,15]. Fourth, we identified an intronic locus within *DLG2* (rs683250; 11q14.1; $n = 26{,}258$; $P = 3.94 \times 10^{-11}$), which encodes the postsynaptic density 93 (PSD-93) protein (Extended Data Fig. 3d). PSD-93 is a membrane-associated guanylate kinase involved in organizing channels in the postsynaptic density[16]. *DLG2* expression increases during early mid-fetal development in the striatum[13] (Extended Data Fig. 8d). Genetic variants in *DLG2* affect learning and cognitive flexibility[17] and are associated with schizophrenia[18]. Notably, SNPs associated with variation in putamen volume showed enrichment of genes involved in apoptosis and axon guidance pathways (Extended Data Fig. 7 and Supplementary Table 7).

Hippocampal volume showed an intergenic association near the *HRK* gene (rs77956314; 12q24.22; $n = 17{,}190$; $P = 2.82 \times 10^{-15}$; Extended Data Fig. 3g) and with an intronic locus in the *MSRB3* gene (rs61921502; 12q14.3; $n = 16{,}209$; $P = 6.87 \times 10^{-11}$; Extended Data Fig. 3h), supporting our previous analyses[5,19] of smaller samples imputed to HapMap3 references. Caudate volume was associated with an intergenic locus 80 kb from *FAT3* (rs1318862; 11q14.3; $n = 15{,}031$; $P = 6.17 \times 10^{-9}$; Extended Data Fig. 3e). This gene encodes a cadherin specifically expressed in the

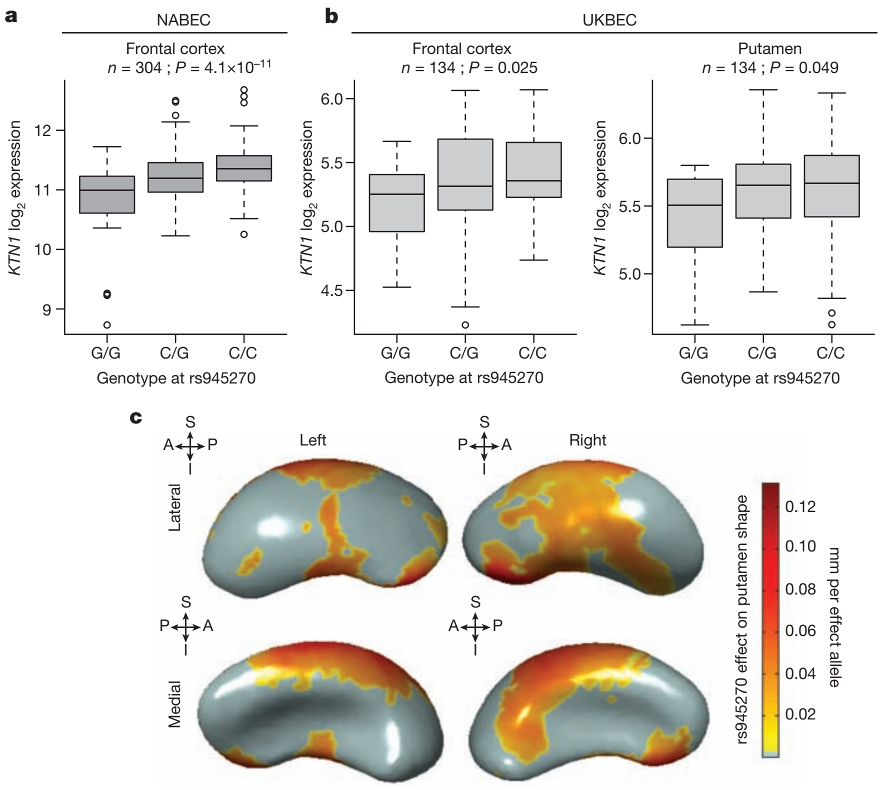


Figure 2 | Effect of rs945270 on *KTN1* expression and putamen shape. **a**, **b**, Expression quantitative trait loci study in brain tissue demonstrates the effect of rs945270 on *KTN1* gene expression in frontal cortex tissue from 304 subjects from the North American Brain Expression Cohort (NABEC[25]) (**a**) and in an independent sample of 134 subjects from the UK Brain Expression Cohort (UKBEC) (**b**), sampled from both frontal cortex and putamen. Boxplot dashed bars mark the twenty-fifth and seventy-fifth percentiles. **c**, Surface-based analysis demonstrates that rs945270 has strong effects on the shape of superior and lateral portions of the putamen in 1,541 subjects. Each copy of the rs945270-C allele was significantly associated with increased width in coloured areas (false discovery rate corrected at $q = 0.05$), and the degree of deformation is labelled by colour, with red indicating greater deformation. Orientation is indicated by arrows. A, anterior; I, inferior; P, posterior, S, superior.

nervous system during embryonic development that influences neuronal morphology through cell–cell interactions[20]. The ICV was associated with an intronic locus within *CRHR1* that tags the chromosome 17q21 inversion[21], which has been previously found to influence ICV[6] (rs17689882; 17q21.31; $n = 12{,}822$; $P = 7.72 \times 10^{-9}$; Extended Data Fig. 3f). Another previously identified variant with association to ICV (rs10784502)[5,19] did not survive genome-wide significance in this analysis but did show a nominal effect in the same direction ($P = 2.05 \times 10^{-3}$; $n = 11{,}373$). None of the genome-wide significant loci in this study were in linkage disequilibrium with known functional coding variants, splice sites, or 3′/5′ untranslated regions, although several of the loci had epigenetic markings suggesting a regulatory role (Extended Data Fig. 3).

Given the strong association with putamen volume, we further examined the rs945270 locus. Epigenetic markers suggest insulator functionality near the locus as this is the lone chromatin mark in the intergenic region[22] (Extended Data Fig. 3a). Chromatin immunoprecipitation followed by sequencing (ChIP-seq) indicate that a variant (rs8017172) in complete linkage disequilibrium with rs945270 ($r^2 = 1.0$) lies within a binding site of the CTCF (CCCTC-binding factor) transcription regulator[23] (Extended Data Fig. 9) in embryonic stem cells. To assess potential functionality in brain tissue, we tested for association with gene expression 1 megabase (Mb) up/downstream. We identified and replicated an effect of rs945270 on the expression of the *KTN1* gene. The C allele, associated with larger putamen volume, also increased expression of *KTN1* in the frontal cortex (discovery sample: 304 neuropathologically normal controls[24] ($P = 4.1 \times 10^{-11}$); replication sample: 134 neuropathologically normal controls ($P = 0.025$)), and putamen (sample: 134 neuropathologically normal controls[25] ($P = 0.049$); Fig. 2a, b). In blood, rs945270 was also strongly associated with *KTN1* expression[26] ($P = 5.94 \times 10^{-31}$; $n = 5{,}311$). After late fetal development, *KTN1* is expressed in the human thalamus, striatum and hippocampus; it is more highly expressed in the striatum than the cortex[13] (Extended Data Fig. 8a). *KTN1* encodes the kinectin receptor facilitating vesicle binding to kinesin, and is heavily involved in organelle transport[11]. Kinectin is only found in the dendrites and soma of neurons, not their axons; neurons with more kinectin have larger cell bodies[27], and kinectin knockdown strongly influences cell shape[28]. The volumetric effects identified here may therefore reflect genetic control of neuronal cell size and/or dendritic complexity. Using three-dimensional surface models of putamen segmentations in MRI scans of 1,541 healthy adolescent subjects, we further localized the allelic effects of rs945270 to regions along the superior and lateral putamen bilaterally, independent of chosen segmentation protocol (Fig. 2c and Extended Data Fig. 10). Each copy of the C allele was associated with an increase in volume along anterior superior regions receiving dense cortical projections from dorsolateral prefrontal cortex and supplementary motor areas[29,30].

In summary, we discovered several common genetic variants underlying variation in different structures within the human brain. Many seem to exert their effects through known developmental pathways including apoptosis, axon guidance and vesicle transport. All structure volumes showed high heritability, but individual genetic variants had diverse effects. The strongest effects were found for putamen and hippocampal volumes, whereas other structures delineated with similar reliability such as the thalamus showed no association with these or other loci (Supplementary Table 4). Discovery of common variants affecting the human brain is now feasible using collaborative analysis of MRI data, and may determine genetic mechanisms driving development and disease.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 April; accepted 19 November 2014.
Published online 21 January 2015.

1. Blokland, G. A., de Zubicaray, G. I., McMahon, K. L. & Wright, M. J. Genetic and environmental influences on neuroimaging phenotypes: a meta-analytical perspective on twin imaging studies. *Twin Res. Hum. Genet.* **15,** 351–371 (2012).
2. Kravitz, A. V. *et al.* Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature* **466,** 622–626 (2010).
3. Poldrack, R. A. *et al.* Interactive memory systems in the human brain. *Nature* **414,** 546–550 (2001).
4. Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J. & Frith, C. D. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442,** 1042–1045 (2006).

5. Stein, J. L. *et al.* Identification of common variants associated with human hippocampal and intracranial volumes. *Nature Genet.* **44,** 552–561 (2012).
6. Ikram, M. A. *et al.* Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nature Genet.* **44,** 539–544 (2012).
7. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467,** 832–838 (2010).
8. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genet.* **42,** 937–948 (2010).
9. van der Sluis, S., Posthuma, D. & Dolan, C. V. TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genet.* **9,** e1003235 (2013).
10. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).
11. Kumar, J., Yu, H. & Sheetz, M. P. Kinectin, an Essential Anchor for Kinesin-Driven Vesicle Motility. *Science* **267,** 1834–1837 (1995).
12. Hamasaki, T., Goto, S., Nishikawa, S. & Ushio, Y. A role of netrin-1 in the formation of the subcortical structure striatum: repulsive action on the migration of late-born striatal neurons. *J. Neurosci.* **21,** 4272–4280 (2001).
13. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478,** 483–489 (2011).
14. Motoyama, N. *et al.* Massive cell death of immature hematopoietic cells and neurons in Bcl-x-deficient mice. *Science* **267,** 1506–1510 (1995).
15. Itoh, K. *et al.* Apoptosis in the basal ganglia of the developing human nervous system. *Acta Neuropathol.* **101,** 92–100 (2001).
16. Scannevin, R. H. & Huganir, R. L. Postsynaptic organization and regulation of excitatory synapses. *Nature Rev. Neurosci.* **1,** 133–141 (2000).
17. Nithianantharajah, J. *et al.* Synaptic scaffold evolution generated components of vertebrate cognitive complexity. *Nature Neurosci.* **16,** 16–24 (2013).
18. Kirov, G. *et al. De novo* CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17,** 142–153 (2012).
19. Bis, J. C. *et al.* Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nature Genet.* **44,** 545–551 (2012).
20. Deans, M. R. *et al.* Control of neuronal morphology by the atypical cadherin Fat3. *Neuron* **71,** 820–832 (2011).
21. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nature Genet.* **37,** 129–137 (2005).
22. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods* **9,** 215–216 (2012).
23. Ziebarth, J. D., Bhattacharya, A. & Cui, Y. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. *Nucleic Acids Res.* **41,** D188–D194 (2013).
24. Hernandez, D. G. *et al.* Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. *Neurobiol. Dis.* **47,** 20–28 (2012).
25. Ramasamy, A. *et al.* Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nature Neurosci.* **17,** 1418–1428 (2014).
26. Westra, H. J. *et al.* Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nature Genet.* **45,** 1238–1243 (2013).
27. Toyoshima, I. & Sheetz, M. P. Kinectin distribution in chicken nervous system. *Neurosci. Lett.* **211,** 171–174 (1996).
28. Zhang, X. *et al.* Kinectin-mediated endoplasmic reticulum dynamics supports focal adhesion growth in the cellular lamella. *J. Cell Sci.* **123,** 3901–3912 (2010).
29. Cohen, M. X., Schoene-Bake, J. C., Elger, C. E. & Weber, B. Connectivity-based segregation of the human striatum predicts personality characteristics. *Nature Neurosci.* **12,** 32–34 (2009).
30. Parent, A. & Hazrati, L. N. Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Res. Brain Res. Rev.* **20,** 91–127 (1995).

Supplementary Information is available in the online version of the paper.

Acknowledgements Funding sources for contributing sites and acknowledgments of contributing consortia authors can be found in Supplementary Note 3.

Author Contributions Individual author contributions are listed in Supplementary Note 4.

Author Information Summary statistics from GWAS results are available online using the ENIGMA-Vis web tool: http://enigma.ini.usc.edu/enigma-vis/. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.M.T. (pthomp@usc.edu) or S.E.M. (Sarah.Medland@qimrberghofer.edu.au).

Derrek P. Hibar[1]*, Jason L. Stein[1,2]*, Miguel E. Renteria[3]*, Alejandro Arias-Vasquez[4,5,6,7]*, Sylvane Desrivières[8]*, Neda Jahanshad[1], Roberto Toro[9,10,11], Katharina Wittfeld[12,13], Lucija Abramovic[14], Micael Andersson[15], Benjamin S. Aribisala[16,17,18], Nicola J. Armstrong[19,20], Manon Bernard[21], Marc M. Bohlken[14], Marco P. Boks[14], Janita Bralten[4,6,7], Andrew A. Brown[22,23], M. Mallar Chakravarty[24,25], Qiang Chen[26], Christopher R. K. Ching[1,27], Gabriel Cuellar-Partida[3], Anouk den Braber[28], Sudheer Giddaluru[29,30], Aaron L. Goldman[26], Oliver Grimm[31], Tulio Guadalupe[32,33], Johanna Hass[34], Girma Woldehawariat[35], Avram J. Holmes[36,37], Martine Hoogman[4,7], Deborah Janowitz[13], Tianye Jia[8], Sungeun Kim[38,39,40], Marieke Klein[4,7], Bernd Kraemer[41], Phil H. Lee[37,42,43,44], Loes M. Olde Loohuis[45], Michelle Luciano[46], Christine Macare[8], Karen A. Mather[19], Manuel Mattheisen[47,48,49], Yuri Milaneschi[50], Kwangsik Nho[38,39,40], Martina Papmeyer[51], Adaikalavan Ramasamy[52,53], Shannon L. Risacher[38,40], Roberto Roiz-Santiañez[54,55], Emma J. Rose[56,57], Alireza Salami[15,58], Philipp G. Sämann[59], Lianne Schmaal[50], Andrew J. Schork[60,61], Jean Shin[21], Lachlan T. Strike[3,62,63], Alexander Teumer[64], Marjolein M. J. van Donkelaar[4,7], Kristel R. van Eijk[14], Raymond K. Walters[65,66], Lars T. Westlye[23,67], Christopher D. Whelan[1], Anderson M. Winkler[68,69], Marcel P. Zwiers[7], Saud Alhusaini[70,71], Lavinia Athanasiu[22,23], Stefan Ehrlich[34,37,72], Marina M. H. Hakobjan[4,7], Cecilie B. Hartberg[22,73], Unn K. Haukvik[22], Angelien J. G. A. M. Heister[4,7], David Hoehn[59], Dalia Kasperaviciute[74,75], David C. M. Liewald[46], Lorna M. Lopez[46], Remco R. R. Makkinje[4,7], Mar Matarin[76], Marlies A. M. Naber[4,7], D. Reese McKay[69,77], Margaret Needham[56], Allison C. Nugent[35], Benno Pütz[59], Natalie A. Royle[16,46,18], Li Shen[38,39,40], Emma Sprooten[51,69,77], Daniah Trabzuni[53,78], Saskia S. L. van der Marel[4,7], Kimm J. E. van Hulzen[4,7], Esther Walton[34], Christiane Wolf[59], Laura Almasy[79,80], David Ames[81,82], Sampath Arepalli[83], Amelia A. Assareh[19], Mark E. Bastin[16,18,46,84], Henry Brodaty[19], Kazima B. Bulayeva[85], Melanie A. Carless[79], Sven Cichon[86,87,88,89], Aiden Corvin[56], Joanne E. Curran[79], Michael Czisch[59], Greig I. de Zubicaray[62], Allissa Dillman[83], Ravi Duggirala[79], Thomas D. Dyer[79,80], Susanne Erk[90], Iryna O. Fedko[28], Luigi Ferrucci[91], Tatiana M. Foroud[40,92], Peter T. Fox[80,93], Masaki Fukunaga[94], J. Raphael Gibbs[53,82], Harald H. H. Göring[79], Robert C. Green[95,96], Sebastian Guelfi[53], Narelle K. Hansell[3], Catharina A. Hartman[97], Katrin Hegenscheid[98], Andreas Heinz[89], Dena G. Hernandez[53,82], Dirk J. Heslenfeld[99], Pieter J. Hoekstra[97], Florian Holsboer[59], Georg Homuth[100], Jouke-Jan Hottenga[28], Masashi Ikeda[101], Clifford R. Jack Jr[102], Mark Jenkinson[103], Robert Johnson[104], Ryota Kanai[105,106], Maria Keil[41], Jack W. Kent Jr[79], Peter Kochunov[107], John B. Kwok[108,109], Stephen M. Lawrie[51], Xinmin Liu[35,110], Dan L. Longo[111], Katie L. McMahon[63], Eva Meisenzahl[112], Ingrid Melle[22,23], Sebastian Mohnke[90], Grant W. Montgomery[3], Jeanette C. Mostert[4,7], Thomas W. Mühleisen[87,88,89], Michael A. Nalls[83], Thomas E. Nichols[103,113], Lars G. Nilsson[15], Markus M. Nöthen[87,89], Kazutaka Ohi[114], Rene L. Olvera[92], Rocio Perez-Iglesias[55,115], G. Bruce Pike[116,117], Steven G. Potkin[118], Ivar Reinvang[67], Simone Reppermund[19], Marcella Rietschel[31], Nina Romanczuk-Seiferth[90], Glenn D. Rosen[119,120], Dan Rujescu[112], Knut Schnell[121], Peter R. Schofield[108,109], Colin Smith[122], Vidar M. Steen[29,30], Jessika E. Sussmann[51], Anbupalam Thalamuthu[19], Arthur W. Toga[123], Bryan J. Traynor[83], Juan Troncoso[124], Jessica A. Turner[125], Maria C. Valdés Hernández[84], Dennis van 't Ent[28], Marcel van der Brug[126], Nic J. A. van der Wee[127], Marie-Jose van Tol[128], Dick J. Veltman[50], Thomas H. Wassink[129], Eric Westman[130], Ronald H. Zielke[104], Alan B. Zonderman[131], David G. Ashbrook[132], Reinmar Hager[132], Lu Lu[133,134,135], Francis J. McMahon[35], Derek W. Morris[56,136], Robert W. Williams[133,134], Han G. Brunner[4,7,137], Randy L. Buckner[37,138], Jan K. Buitelaar[6,7,139], Wiepke Cahn[14], Vince D. Calhoun[140,141], Gianpiero L. Cavalleri[71], Benedicto Crespo-Facorro[54,55], Anders M. Dale[142,143], Gareth E. Davies[144], Norman Delanty[71,145], Chantal Depondt[146], Srdjan Djurovic[22,147], Wayne C. Drevets[35,148], Thomas Espeseth[23,67], Randy L. Gollub[37,72,96], Beng-Choon Ho[149], Wolfgang Hoffmann[12,64], Norbert Hosten[98], René S. Kahn[14], Stephanie Le Hellard[29,30], Andreas Meyer-Lindenberg[31], Bertram Müller-Myhsok[59,150,151], Matthias Nauck[152], Lars Nyberg[15], Massimo Pandolfo[146], Brenda W. J. H. Penninx[50], Joshua L. Roffman[37], Sanjay M. Sisodiya[74], Jordan W. Smoller[37,42,43,96], Hans van Bokhoven[4,7], Neeltje E. M. van Haren[14], Henry Völzke[64], Henrik Walter[90], Michael W. Weiner[153], Wei Wen[19], Tonya White[154,155], Ingrid Agartz[22,73,156], Ole A. Andreassen[22,23], John Blangero[79,80], Dorret I. Boomsma[28], Rachel M. Brouwer[14], Dara M. Cannon[35,157], Mark R. Cookson[83], Eco J. C. de Geus[28], Ian J. Deary[46], Gary Donohoe[56,136], Guillén Fernández[6,7], Simon E. Fisher[7,32], Clyde Francks[7,32], David C. Glahn[69,77], Hans J. Grabe[13,158], Oliver Gruber[41,59], John Hardy[53], Ryota Hashimoto[159], Hilleke E. Hulshoff Pol[14], Erik G. Jönsson[22,156], Iwona Kloszewska[160], Simon Lovestone[161,162], Venkata S. Mattay[26,163], Patrizia Mecocci[164], Colm McDonald[157], Andrew M. McIntosh[46,51], Roel A. Ophoff[14,45], Tomas Paus[165,166], Zdenka Pausova[21,167], Mina Ryten[53,52], Perminder S. Sachdev[19,168], Andrew J. Saykin[38,40,90], Andy Simmons[169,170,171], Andrew Singleton[83], Hilkka Soininen[172,173], Joanna M. Wardlaw[16,18,46,84], Michael E. Weale[52], Daniel R. Weinberger[26,174], Hieab H. H. Adams[155,175], Lenore J. Launer[176], Stephan Seiler[177], Reinhold Schmidt[177], Ganesh Chauhan[178], Claudia L. Satizabal[179,180], James T. Becker[181,182,183], Lisa Yanek[184], Sven J. van der Lee[175], Maritza Ebling[72,185], Bruce Fischl[72,185,186], W. T. Longstreth Jr[187], Douglas Greve[72,185], Helena Schmidt[188], Paul Nyquist[189], Louis N. Vinke[72,185], Cornelia M. van Duijn[175], Luting Xue[190], Bernard Mazoyer[191], Joshua C. Bis[192], Vilmundur Gudnason[193], Sudha Seshadri[179,181], M. Arfan Ikram[155,175], The Alzheimer's Disease Neuroimaging Initiative†, The CHARGE Consortium†, EPIGEN†, IMAGEN†, SYS†, Nicholas G. Martin[3]§, Margaret J. Wright[3,62]§, Gunter Schumann[8]§, Barbara Franke[4,5,7]§, Paul M. Thompson[1]§ & Sarah E. Medland[3]§

[1]Imaging Genetics Center, Institute for Neuroimaging & Informatics, Keck School of Medicine of the University of Southern California, Los Angeles, California 90292, USA. [2]Neurogenetics Program, Department of Neurology, UCLA School of Medicine, Los Angeles, California 90095, USA. [3]QIMR Berghofer Medical Research Institute, Brisbane 4006, Australia. [4]Department of Human Genetics, Radboud university medical center, Nijmegen 6500 HB, The Netherlands. [5]Department of Psychiatry, Radboud university medical center, Nijmegen 6500 HB, The Netherlands. [6]Department of Cognitive Neuroscience, Radboud university medical center, Nijmegen 6500 HB, The Netherlands. [7]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen 6500 GL, The Netherlands. [8]MRC-SGDP Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, UK. [9]Laboratory of Human Genetics and Cognitive Functions, Institut Pasteur, Paris 75015, France. [10]Centre Nationale de Recherche Scientifique (CNRS) Unité de Recherche Associée (URA) 2182 Genes, Synapses and Cognition, Institut Pasteur, Paris 75015, France. [11]Université Paris

Diderot, Sorbonne Paris Cité, Paris 75015, France. [12]German Center for Neurodegenerative Diseases (DZNE) Rostock/Greifswald, Greifswald 17487, Germany. [13]Department of Psychiatry, University Medicine Greifswald, Greifswald 17489, Germany. [14]Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands. [15]Umeå Centre for Functional Brain Imaging (UFBI), Umeå University, Umeå 901 87, Sweden. [16]Brain Research Imaging Centre, University of Edinburgh, Edinburgh EH4 2XU, UK. [17]Department of Computer Science, Lagos State University, Lagos, Nigeria. [18]Scottish Imaging Network, A Platform for Scientific Excellence (SINAPSE) Collaboration, Department of Neuroimaging Sciences, University of Edinburgh, Edinburgh EH4 2XU, UK. [19]Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney 2052, Australia. [20]School of Mathematics and Statistics, University of Sydney, Sydney 2006, Australia. [21]The Hospital for Sick Children, University of Toronto, Toronto M5G 1X8, Canada. [22]NORMENT - KG Jebsen Centre, Institute of Clinical Medicine, University of Oslo, Oslo N-0316, Norway. [23]NORMENT - KG Jebsen Centre, Division of Mental Health and Addiction, Oslo University Hospital, Oslo 0424, Norway. [24]Cerebral Imaging Centre, Douglas Mental Health University Institute, Montreal H4H 1R3, Canada. [25]Department of Psychiatry and Biomedical Engineering, McGill University, Montreal H3A 2B4, Canada. [26]Lieber Institute for Brain Development, Baltimore, Maryland 21205, USA. [27]Interdepartmental Neuroscience Graduate Program, UCLA School of Medicine, Los Angeles, California 90095, USA. [28]Biological Psychology, Neuroscience Campus Amsterdam & EMGO Institute for Health and Care Research, VU University & VU Medical Center, Amsterdam 1081 BT, The Netherlands. [29]NORMENT - KG Jebsen Centre for Psychosis Research, Department of Clinical Science, University of Bergen, 5021 Bergen, Norway. [30]Dr. Einar Martens Research Group for Biological Psychiatry, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen 5021, Norway. [31]Central Institute of Mental Health, Medical Faculty Mannheim, University Heidelberg, Mannheim 68159, Germany. [32]Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen 6525 XD, The Netherlands. [33]International Max Planck Research School for Language Sciences, Nijmegen 6525 XD, The Netherlands. [34]Department of Child and Adolescent Psychiatry, Faculty of Medicine of the TU Dresden, Dresden 01307 Germany. [35]Human Genetics Branch and Experimental Therapeutics and Pathophysiology Branch, National Institute of Mental Health Intramural Research Program, Bethesda, Maryland 20892, USA. [36]Department of Psychology, Yale University, New Haven, Connecticut 06511, USA. [37]Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02115, USA. [38]Center for Neuroimaging, Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [39]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [40]Indiana Alzheimer Disease Center, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [41]Center for Translational Research in Systems Neuroscience and Psychiatry, Department of Psychiatry and Psychotherapy, University Medical Center, Goettingen 37075, Germany. [42]Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02115, USA. [43]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Boston, Massachusetts 02141, USA. [44]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts 02115, USA. [45]Center for Neurobehavioral Genetics, University of California, Los Angeles, California 90095, USA. [46]Centre for Cognitive Ageing and Cognitive Epidemiology, Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK. [47]Department of Biomedicine, Aarhus University, Aarhus DK-8000, Denmark. [48]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus and Copenhagen DK-8000, Denmark. [49]Center for integrated Sequencing, iSEQ, Aarhus University, Aarhus DK-8000, Denmark. [50]Department of Psychiatry, Neuroscience Campus Amsterdam, VU University Medical Center/GGZ inGeest, Amsterdam 1081 HL, The Netherlands. [51]Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, Edinburgh EH10 5HF, UK. [52]Department of Medical and Molecular Genetics, King's College London, London SE1 9RT, UK. [53]Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, London WC1N 3BG, UK. [54]Department of Psychiatry, University Hospital Marqués de Valdecilla, School of Medicine, University of Cantabria-IDIVAL, Santander 39008, Spain. [55]Cibersam (Centro Investigación Biomédica en Red Salud Mental), Madrid 28029, Spain. [56]Neuropsychiatric Genetics Research Group and Department of Psychiatry, Trinity College Institute of Psychiatry, Trinity College Dublin, Dublin 2, Ireland. [57]Center for Translational Research on Adversity, Neurodevelopment and Substance Abuse (C-TRANS), Department of Psychiatry, University of Maryland School of Medicine, Baltimore, Maryland 21045, USA. [58]Aging Research Center, Karolinska Institutet and Stockholm University, 11330 Stockholm, Sweden. [59]Max Planck Institute of Psychiatry, Munich 80804, Germany. [60]Multimodal Imaging Laboratory, Department of Neurosciences, University of California, San Diego, California 92093, USA. [61]Department of Cognitive Sciences, University of California, San Diego, California 92161, USA. [62]School of Psychology, University of Queensland, Brisbane 4072, Australia. [63]Centre for Advanced Imaging, University of Queensland, Brisbane 4072, Australia. [64]Institute for Community Medicine, University Medicine Greifswald, Greifswald D-17475, Germany. [65]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [66]Medical and Population Genetics Program, Broad Institute of Harvard and MIT, Boston, Massachusetts 02142, USA. [67]Department of Psychology, University of Oslo, Oslo 0373, Norway. [68]The Oxford Centre for Functional MRI of the Brain, Nuffield Department of Clinical Neurosciences, Oxford University, Oxford OX3 9DU, UK. [69]Department of Psychiatry, Yale School of Medicine, New Haven, Connecticut 06511, USA. [70]Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal H3A 2B4, Canada. [71]Molecular and Cellular Therapeutics, The Royal College of Surgeons, Dublin 2, Ireland. [72]The Athinoula A.Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. [73]Department of Psychiatric Research and Development, Diakonhjemmet Hospital, Oslo 0319, Norway. [74]UCL Institute of Neurology, London, United Kingdom and Epilepsy Society, London WC1N 3BG, UK. [75]Department of Medicine, Imperial College London, London W12 0NN, UK. [76]Department of Clinical and Experimental Epilepsy, UCL Institute of Neurology, London WC1N 3BG, UK. [77]Olin Neuropsychiatric Research Center, Institute of Living, Hartford Hospital, Hartford, Connecticut 06106, USA. [78]Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh 11211, Saudi Arabia. [79]Texas Biomedical Research Institute, San Antonio, Texas 78245, USA. [80]University of Texas Health Science Center, San Antonio, Texas 78229, USA. [81]National Ageing Research Institute, Royal Melbourne Hospital, Melbourne 3052, Australia. [82]Academic Unit for Psychiatry of Old Age, University of Melbourne, Melbourne 3101, Australia. [83]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. [84]Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh EH4 2XU, UK. [85]N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow 119333, Russia. [86]Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel 4055, Switzerland. [87]Institute of Human Genetics, University of Bonn, Bonn, D-53127, Germany. [88]Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Jülich, D-52425, Germany. [89]Department of Genomics, Life & Brain Center, University of Bonn, Bonn D-53127, Germany. [90]Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin Berlin, CCM, Berlin 10117, Germany. [91]Clinical Research Branch, National Institute on Aging, Baltimore, Maryland 20892, USA. [92]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [93]South Texas Veterans Health Care System, San Antonio, Texas 78229, USA. [94]Biofunctional Imaging, Immunology Frontier Research Center, Osaka University, Osaka 565-0871, Japan. [95]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [96]Harvard Medical School, Boston, Massachusetts 02115, USA. [97]Department of Psychiatry, University of Groningen, University Medical Center Groningen, 9713 GZ Groningen, The Netherlands. [98]Institute of Diagnostic Radiology and Neuroradiology, University Medicine Greifswald, Greifswald 17475, Germany. [99]Departments of Cognitive and Clinical Neuropsychology, VU University Amsterdam, 1081 BT Amsterdam, The Netherlands [100]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald 17489, Germany. [101]Department of Psychiatry, Fujita Health University School of Medicine, Toyoake 470-1192, Japan. [102]Radiology, Mayo Clinic, Rochester, Minnesota 55905, USA. [103]FMRIB Centre, University of Oxford, Oxford OX3 9DU, UK. [104]NICHD Brain and Tissue Bank for Developmental Disorders, University of Maryland Medical School, Baltimore, Maryland 21201, USA. [105]School of Psychology, University of Sussex, Brighton BN1 9QH, UK. [106]Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, UK. [107]Department of Psychiatry, Maryland Psychiatric Research Center, University of Maryland, Baltimore, Maryland 21201, USA. [108]Neuroscience Research Australia, Sydney 2031, Australia. [109]School of Medical Sciences, UNSW, Sydney 2052, Australia. [110]Department of Pathology and Cell Biology, Columbia University Medical Center, New York 10032, USA. [111]Lymphocyte Cell Biology Unit, Laboratory of Genetics, National Institute on Aging, National Institutes of Health, Baltimore, Maryland 21224, USA. [112]Department of Psychiatry, Ludwig-Maximilians-Universität, Munich 80336, Germany. [113]Department of Statistics & WMG, University of Warwick, Coventry CV4 7AL, UK. [114]Department of Psychiatry, Osaka University Graduate School of Medicine, Osaka 565-0871, Japan. [115]Institute of Psychiatry, King's College London, London SE5 8AF, UK. [116]Department of Neurology, University of Calgary, Calgary T2N 2T9, Canada. [117]Department of Clinical Neuroscience, University of Calgary, Calgary T2N 2T9, Canada. [118]Psychiatry and Human Behavior, University of California, Irvine, California 92617, USA. [119]Department of Neurology, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. [120]Harvard Medical School, Boston, Massachusetts 02115, USA. [121]Department of General Psychiatry, Heidelberg University Hospital, Heidelberg 69115, Germany. [122]Department of Neuropathology, MRC Sudden Death Brain Bank Project, University of Edinburgh, Edinburgh EH8 9AG, UK. [123]Laboratory of Neuro Imaging, Institute for Neuroimaging and Informatics, Keck School of Medicine of the University of Southern California, Los Angeles, California 90033, USA. [124]Department of Pathology, Johns Hopkins University, Baltimore, Maryland 21287, USA. [125]Psychology Department and Neuroscience Institute, Georgia State University, Atlanta, Georgia 30302, USA. [126]Genentech, South San Francisco, California 94080, USA [127]Psychiatry and Leiden Institute for Brain and Cognition, Leiden University Medical Center, Leiden 2333 ZA, The Netherlands. [128]Neuroimaging Centre, University of Groningen, University Medical Center Groningen, Groningen 9713 AW, The Netherlands. [129]Department of Psychiatry, Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA. [130]Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm SE-141 83, Sweden. [131]Behavioral Epidemiology Section, National Institute on Aging Intramural Research Program, Baltimore, Maryland 20892, USA. [132]Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK. [133]Center for Integrative and Translational Genomics, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA. [134]Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA. [135]Jiangsu Province Key Laboratory for Inflammation and Molecular Drug Target, Medical College of Nantong University, Nantong 226001, China. [136]Cognitive Genetics and Therapy Group, School of Psychology & Discipline of Biochemistry, National University of Ireland Galway, Galway, Ireland. [137]Department of Clinical Genetics, Maastricht University Medical Center, Maastricht 6200 MD, The Netherlands. [138]Department of Psychology, Center for Brain Science, Harvard University, Boston, Massachusetts 02138, USA. [139]Karakter Child and Adolescent Psychiatry, Radboud university medical center, Nijmegen 6500 HB, The Netherlands. [140]The Mind Research Network & LBERI, Albuquerque, New Mexico 87106, USA. [141]Department of ECE, University of New Mexico, Albuquerque, New Mexico 87131, USA. [142]Center for Translational Imaging and Personalized Medicine, University of California, San Diego, California 92093, USA. [143]Departments of Neurosciences, Radiology, Psychiatry, and Cognitive Science, University of California, San Diego, California 92093, USA. [144]Avera Institute for Human Genetics, Sioux Falls, South Dakota, 57108, USA. [145]Neurology Division, Beaumont Hospital, Dublin 9, Ireland. [146]Department of Neurology, Hopital Erasme, Universite Libre de Bruxelles, Brussels 1070, Belgium. [147]Department of Medical Genetics, Oslo University Hospital, Oslo 0450, Norway. [148]Janssen Research & Development, Johnson & Johnson, Titusville, New Jersey 08560, USA. [149]Department of

Psychiatry, University of Iowa, Iowa City, Iowa 52242, USA. [150]Munich Cluster for Systems Neurology (SyNergy), Munich 81377, Germany. [151]University of Liverpool, Institute of Translational Medicine, Liverpool L69 3BX, UK. [152]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald 17475, Germany. [153]Center for Imaging of Neurodegenerative Disease, San Francisco VA Medical Center, University of California, San Francisco, California 94121, USA. [154]Department of Child and Adolescent Psychiatry, Erasmus University Medical Centre, Rotterdam 3000 CB, The Netherlands. [155]Department of Radiology, Erasmus University Medical Centre, Rotterdam 3015 CN, The Netherlands. [156]Department of Clinical Neuroscience, Psychiatry Section, Karolinska Institutet, Stockholm SE-171 76, Sweden. [157]Clinical Neuroimaging Laboratory, College of Medicine, Nursing and Health Sciences, National University of Ireland Galway, Galway, Ireland. [158]Department of Psychiatry and Psychotherapy, HELIOS Hospital Stralsund 18435, Germany. [159]Molecular Research Center for Children's Mental Development, United Graduate School of Child Development, Osaka University, Osaka 565-0871, Japan. [160]Medical University of Lodz, Lodz 90-419, Poland. [161]Department of Psychiatry, University of Oxford, Oxford OX3 7JX, UK. [162]NIHR Dementia Biomedical Research Unit, King's College London, London SE5 8AF, UK. [163]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [164]Section of Gerontology and Geriatrics, Department of Medicine, University of Perugia, Perugia 06156, Italy. [165]Rotman Research Institute, University of Toronto, Toronto M6A 2E1, Canada. [166]Departments of Psychology and Psychiatry, University of Toronto, Toronto M5T 1R8, Canada. [167]Departments of Physiology and Nutritional Sciences, University of Toronto, Toronto M5S 3E2, Canada. [168]Neuropsychiatric Institute, Prince of Wales Hospital, Sydney 2031, Australia. [169]Department of Neuroimaging, Institute of Psychiatry, King's College London, London SE5 8AF, UK. [170]Biomedical Research Centre for Mental Health, King's College London, London SE5 8AF, UK. [171]Biomedical Research Unit for Dementia, King's College London, London SE5 8AF, UK. [172]Institute of Clinical Medicine, Neurology, University of Eastern Finland, Kuopio FI-70211, Finland. [173]Neurocentre Neurology, Kuopio University Hospital, Kuopio FI-70211, Finland. [174]Departments of Psychiatry, Neurology, Neuroscience and the Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [175]Department of Epidemiology, Erasmus University Medical Centre, Rotterdam 3015 CN, The Netherlands. [176]Laboratory of Epidemiology and Population Sciences, Intramural Research Program, National Institute on Aging, Bethesda, Maryland 20892, USA. [177]Department of Neurology, Clinical Division of Neurogeriatrics, Medical University Graz, Graz 8010, Austria. [178]INSERM U897, University of Bordeaux, Bordeaux 33076, France. [179]Department of Neurology, Boston University School of Medicine, Boston, Massachusetts 02118, USA. [180]Framingham Heart Study, Framingham, Massachusetts 01702, USA. [181]Department of Neurology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. [182]Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. [183]Department of Psychology, Dietrich School of Arts and Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. [184]General Internal Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA. [185]Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. [186]Computer Science and AI Lab, Massachusetts Institute of Technology, Boston, Massachusetts 02141, USA. [187]Department of Neurology University of Washington, Seattle, Washington 98195, USA. [188]Institute of Molecular Biology and Biochemistry, Medical University Graz, 8010 Graz, Austria. [189]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [190]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118, USA. [191]Groupe d'Imagerie Neurofonctionnelle, UMR5296 CNRS, CEA and University of Bordeaux, Bordeaux 33076, France. [192]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington 98101, USA. [193]Icelandic Heart Association, University of Iceland, Faculty of Medicine, Reykjavik 101, Iceland.

†A list of authors and affiliations appears in the Supplementary Information.
*These authors contributed equally to this work.
§These authors jointly supervised this work.

METHODS

Details of the GWAS meta-analysis are outlined in Extended Data Fig. 1. All participants in all cohorts in this study gave written informed consent and sites involved obtained approval from local research ethics committees or Institutional Review Boards. The ENIGMA consortium follows a rolling meta-analysis framework for incorporating sites into the analysis. The discovery sample comprises studies of European ancestry (Extended Data Fig. 2) that contributed GWAS summary statistics for the purpose of this analysis on or before 1 October 2013. The deadline for discovery samples to upload their data was made before inspecting the data and was not influenced by the results of the analyses. The meta-analysed results from discovery cohorts were carried forward for secondary analyses and functional validation studies. Additional samples of European ancestry were gathered to provide *in silico* or single genotype replication of the strongest associations as part of the replication sample. A generalization sample of sites with non-European ancestry was used to examine the effects across ethnicities. In all, data were contributed from 50 cohorts, each of which is detailed in Supplementary Tables 1–3.

The brain measures examined in this study were obtained from structural MRI data collected at participating sites around the world. Brain scans were processed and examined at each site locally, following a standardized protocol procedure to harmonize the analysis across sites. The standardized protocols for image analysis and quality assurance are openly available online (http://enigma.ini.usc.edu/protocols/imaging-protocols/). The subcortical brain measures (nucleus accumbens, amygdala, caudate nucleus, hippocampus, pallidum, putamen and thalamus) were delineated in the brain using well-validated, freely available brain segmentation software packages: FIRST[31], part of the FMRIB Software Library (FSL), or FreeSurfer[32]. The agreement between the two software packages has been well documented in the literature[5,33] and was further detailed here (Supplementary Table 4). Participating sites used the software package most suitable for their data set (the software used at each site is given in Supplementary Table 2) without selection based on genotype or the associations present in this study. In addition to the subcortical structures of the brain, we examined the genetic effects of a measure of global head size, the ICV. The ICV was calculated as: 1/(determinant of a rotation-translation matrix obtained after affine registration to a common study template and multiplied by the template volume (1,948,105 mm^3)). After image processing, each image was inspected individually to identify poorly segmented structures. Each site contributed histograms of the distribution of volumes for the left and right hemisphere structures (and a measure of asymmetry) of each subcortical region used in the analysis. Scans marked as outliers (>3 standard deviations from the mean) based on the histogram plots were re-checked at each site to locate any errors. If a scan had an outlier for a given structure, but was segmented properly, it was retained in the analysis. Site-specific phenotype histograms, Manhattan plots and quantile–quantile plots from each participating site are available on the ENIGMA website (http://enigma.ini.usc.edu/publications/enigma-2/).

Each study in the discovery sample was genotyped using commercially available platforms. Before imputation, genetic homogeneity was assessed in each sample using multi-dimensional scaling (MDS) analysis (Extended Data Fig. 2). Ancestry outliers were excluded through visual inspection of the first two components. Quality control filtering was applied to remove genotyped SNPs with low minor allele frequency (<0.01), poor genotype call rate (<95%), and deviations from Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$) before imputation. The imputation protocols used MaCH[34] for haplotype phasing and minimac[35] for imputation and are freely available online (http://enigma.ini.usc.edu/protocols/genetics-protocols/). Full details of quality control procedures and any deviations from the imputation protocol are given in Supplementary Table 3.

Genome-wide association scans were conducted at each site for all eight traits of interest including the ICV and bilateral volumes of the nucleus accumbens, amygdala, caudate nucleus, hippocampus, pallidum, putamen and thalamus. For each SNP in the genome, the additive dosage value was regressed against the trait of interest separately using a multiple linear regression framework controlling for age, age^2, sex, 4 MDS components, ICV (for non-ICV phenotypes) and diagnosis (when applicable). For studies with data collected from several centres or scanners, dummy-coded covariates were also included in the model. Sites with family data (NTR-Adults, BrainSCALE, QTIM, SYS, GOBS, ASPSFam, ERF, GeneSTAR, NeuroIMAGE and OATS) used mixed-effects models to control for familial relationships in addition to covariates stated previously. The primary analyses for this paper focused on the full set of subjects including data sets with patients to maximize the power to detect effects. We re-analysed the data excluding patients to verify that detected effects were not due to disease alone (Extended Data Fig. 5a). The protocols used for testing association with mach2qtl (ref. 34) for studies with unrelated subjects and merlin-offline[36] for family-based designs are freely available online (http://enigma.ini.usc.edu/protocols/genetics-protocols/). Full details for the software used at each site are given in Supplementary Table 3.

The GWAS results from each site were uploaded to a centralized server for quality checking and processing. Results files from each cohort were free from genomic inflation in quantile–quantile plots and Manhattan plots (http://enigma.ini.usc.edu/publications/enigma-2/). Poorly imputed SNPs (with $R^2 < 0.5$) and low minor allele count (<10) were removed from the GWAS result files from each site. The resulting files were combined meta-analytically using a fixed-effect, inverse-variance-weighted model as implemented in the software package METAL[37]. The discovery cohorts were meta-analysed first, controlling for genomic inflation. The combined discovery data set (comprised of all meta-analysed SNPs with data from at least 5,000 subjects) was carried forward for the additional analyses detailed below.

To account appropriately for multiple comparisons over the eight traits in our analysis, we first examined the degree of independence between each trait. We generated an 8×8 correlation matrix based on the Pearson's correlation between all pair-wise combinations of the mean volumes of each structure in the QTIM study. Using the matSpD software[38] we found that the effective number of independent traits in our analysis was 7. We therefore set a significance criteria threshold of $P < (5 \times 10^{-8}/7) = 7.1 \times 10^{-9}$.

Heritability estimates for mean volumes of each of the eight structures in this study were calculated using structural equation modelling in OpenMx[39]. Twin modelling was performed controlling for age and sex differences on a large sample ($n = 1{,}030$) of healthy adolescent and young adult twins (148 monozygotic and 202 dizygotic pairs) and their siblings from the Queensland Twin Imaging (QTIM) study. Subsequently, a multivariate analysis showed that common environmental factors (C) could be dropped from the model without a significant reduction in the goodness-of-fit ($\Delta\chi^2_{36} = 29.81$; $P = 0.76$). Heritability (h^2) was significantly different from zero for all eight brain measures: putamen ($h^2 = 0.89$; 95% confidence interval 0.85–0.92), thalamus ($h^2 = 0.88$; 0.85–0.92), ICV ($h^2 = 0.88$; 0.84–0.90), hippocampus ($h^2 = 0.79$; 0.74–0.83), caudate nucleus ($h^2 = 0.78$; 0.75–0.82), pallidum ($h^2 = 0.75$; 0.72–0.78), nucleus accumbens ($h^2 = 0.49$; 0.45–0.55), amygdala ($h^2 = 0.43$; 0.39, 0.48) (Extended Data Fig. 11a).

Percentage variance explained by each genome-wide significant SNP was determined based on the final combined discovery data set (Extended Data Fig. 6a) or the discovery combined with the replication samples (Table 1) after correction for covariates using the following equation:

$$R^2_{g|c}/(1-R^2_c) = (t^2/((n-k-1)+t^2)) * 100$$

where the t-statistic is calculated as the beta coefficient for a given SNP from the regression model (controlling for covariates) divided by the standard error of the beta estimate, and where n is the total number of subjects and k is the total number of covariates included in the model ($k = 10$) (ref. 40). $R^2_{g|c}$ is the variance explained by the variant controlling for covariates and R^2_c is the variance explained by the covariates alone. $R^2_{g|c}/(1 - R^2_c)$ gives the variance explained by the genetic variant after accounting for covariate effects. The total variance explained by the GWAS (Extended Data Fig. 11b, c) was calculated by first linkage disequilibrium pruning the results without regard to significance (pruning parameters in PLINK:––indep-pairwise 1000kb 25 0.1). The t-statistics of the regression coefficients from the pruned results are then corrected for the effects of 'winner's curse' and the variance explained by each SNP after accounting for covariate effects is summed across SNPs using freely available code (http://sites.google.com/site/honcheongso/software/total-vg)[40,41]. As the correction for winners curse may be influenced by asymmetry in the distribution of t (arising from the choice of reference allele) we bootstrapped the choice of reference allele (5,000 iterations) to derive the median value and 95% confidence intervals of the estimates of variance explained (Extended Data Fig. 11b, c). The correction for winner's curse corrected for upward biases when estimating the percentage variance explained by each SNP across the genome via simulation[40], but this correction could still allow some bias. Future large studies will be able to evaluate independently the percentage variance explained.

We performed multivariate GWAS using the Trait-based Association Test that uses Extended Simes procedure (TATES)[9]. For the TATES analysis we used GWAS summary statistics from the discovery data set and the correlation matrix created from the eight phenotypes using the QTIM data set (Extended Data Fig. 6c).

We examined the moderating effects of mean age and proportion of females on the effect sizes estimated for the top loci influencing brain volumes (Extended Data Fig. 5b, c) using a mixed-effect meta-regression model such that:

$$\text{effect} = \beta_0 + \beta_{\text{mod}} X_{\text{mod}} + \varepsilon + \eta$$

In this model, the effect and variance at each site are treated as random effects and the moderator X_{mod} (either mean age or proportion of females) is treated as a fixed effect. Meta-regression tests were performed using the metafor package (version 1.9-1) in R.

Hierarchical clustering was performed on the GWAS t-statistics from the discovery data set results using independent SNPs clumped from the TATES results (clumping parameters: significance threshold for index SNP = 0.01, significance

threshold for clumped SNPs = 0.01, $r^2 = 0.25$, physical distance = 1 Mb; Extended Data Fig. 6b). Regions with the strongest genetic similarity were grouped together based on the strength of their pairwise correlations. The results were represented visually using hierarchical clustering with default settings from the gplots package (version 2.12.1) in R.

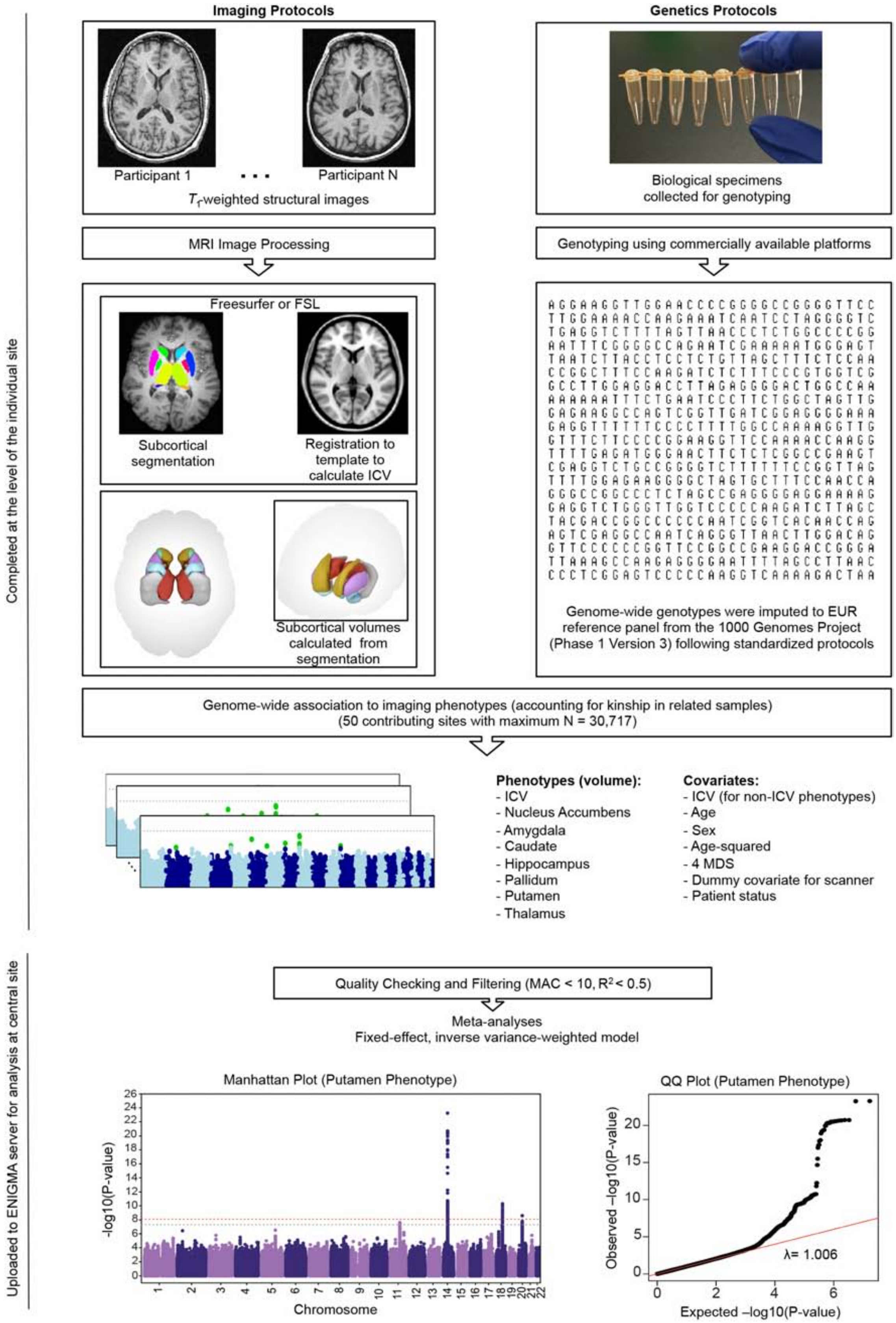
Gene annotation, gene-based test statistics and pathway analysis were performed using the KGG2.5 software package[42] (Supplementary Table 7 and Extended Data Fig. 7). Linkage disequilibrium was calculated based on RSID numbers using the 1000 Genomes Project European samples as a reference (http://enigma.ini.usc.edu/protocols/genetics-protocols/). For the annotation, SNPs were considered 'within' a gene if they fell within 5 kb of the 3′/5′ untranslated regions based on human genome (hg19) coordinates. Gene-based tests were performed using the GATES test[42] without weighting P values by predicted functional relevance. Pathway analysis was performed using the hybrid set-based test (HYST) of association[43]. For all gene-based tests and pathway analyses, results were considered significant if they exceeded a Bonferroni correction threshold accounting for the number of pathways and traits tested such that $P_{\text{thresh}} = 0.05/(671 \text{ pathways} \times 7 \text{ independent traits}) = 1.06 \times 10^{-5}$.

Expression quantitative loci were examined in two independent data sets: the NABEC (GSE36192)[24] and UKBEC (GSE46706)[44,45]. Detailed processing and exclusion criteria for both data sets are described elsewhere[24,45]. In brief, the UKBEC consists of 134 neuropathologically normal donors from the MRC Sudden Death Brain Bank in Edinburgh and Sun Health Research Institute; expression was profiled on the Affymetrix Exon 1.0 ST array. The NABEC is comprised of 304 neurologically normal donors from the National Institute of Ageing and expression profiled on the Illumina HT12v3 array. The expression values were corrected for gender and batch effects and probes that contained polymorphisms (seen >1% in European 1000G) were excluded from analyses[44]. Blood expression quantitative trait loci (eQTL) data were queried using the Blood eQTL Browser (http://genenetwork.nl/bloodeqtlbrowser/)[26]. Brain expression over the lifespan was measured from a spatio-temporal atlas of human gene expression and graphed using custom R scripts (GSE25219; details given in[13]).

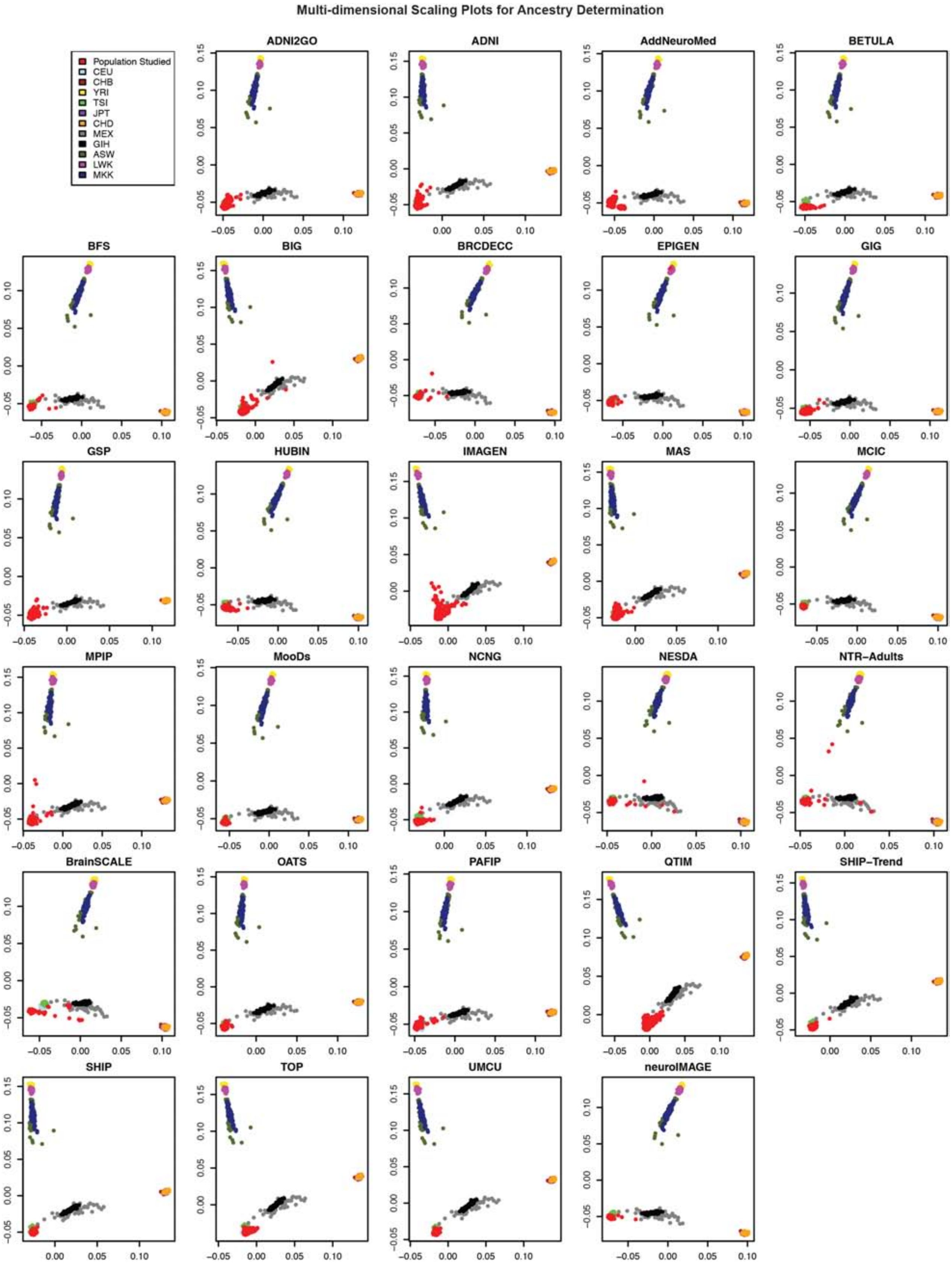
Fine-grained three-dimensional surface mappings of the putamen were generated using a medial surface modelling method[46,47] in 1,541 healthy subjects from the IMAGEN study[48] (Fig. 2c and Extended Data Fig. 10a, b). Putamen volume segmentations from either FSL (Fig. 2c and Extended Data Fig. 10a) or FreeSurfer (Extended Data Fig. 10b) were first converted to three-dimensional meshes and then co-registered to an average template for statistical analysis. The medial core distance was used as a measure of shape and was calculated as the distance from each point on the surface to the centre of the putamen. At each point along the surface of the putamen, an association test was performed using multiple linear regression in which the medial core distance at a given point on the surface was the outcome measure and the additive dosage value of the top SNP was the predictor of interest while including the same covariates that were used for volume including age, sex, age^2, 4 MDS, ICV and site.

In Extended Data Fig. 3, all tracks were taken from the UCSC Genome Browser Human hg19 assembly. SNPs (top 5%) shows the top 5% associated SNPs within the locus and are coloured by their correlation to the top SNP. Genes shows the gene models from GENCODE version 19. Conservation was defined at each base through the phyloP algorithm which assigns scores as $-\log_{10} P$ values under a null hypothesis of neutral evolution calculated from pre-computed genomic alignment of 100 vertebrate species[49]. Conserved sites are assigned positive scores, while faster-than-neutral evolving sites are given negative scores. TFBS conserved shows computationally predicted transcription factor binding sites using the Transfac Matrix Database (v.7.0) found in human, mouse and rat. Brain histone (1.3 year) and brain histone (68 year) show maps of histone trimethylation at histone H3 Lys 4 (H3K4me3), an epigenetic mark for transcriptional activation, measured by ChIP-seq. These measurements were made in neuronal nuclei (NeuN+) collected from prefrontal cortex of post-mortem human brain[50]. CpG methylation was generated using methylated DNA immunoprecipitation and sequencing from postmortem human frontal cortex of a 57-year-old male[51]. DNaseI hypersens displays DNaseI hypersensitivity, evidence of open chromatin, which was evaluated in postmortem human frontal cerebrum from three donors (age 22–35), through the ENCODE consortium[52]. Finally, hES Chrom State gives the predicted chromatin states based on computational integration of ChIP-seq data for nine chromatin marks in H1 human embryonic stem cell lines derived in the ENCODE consortium[53].

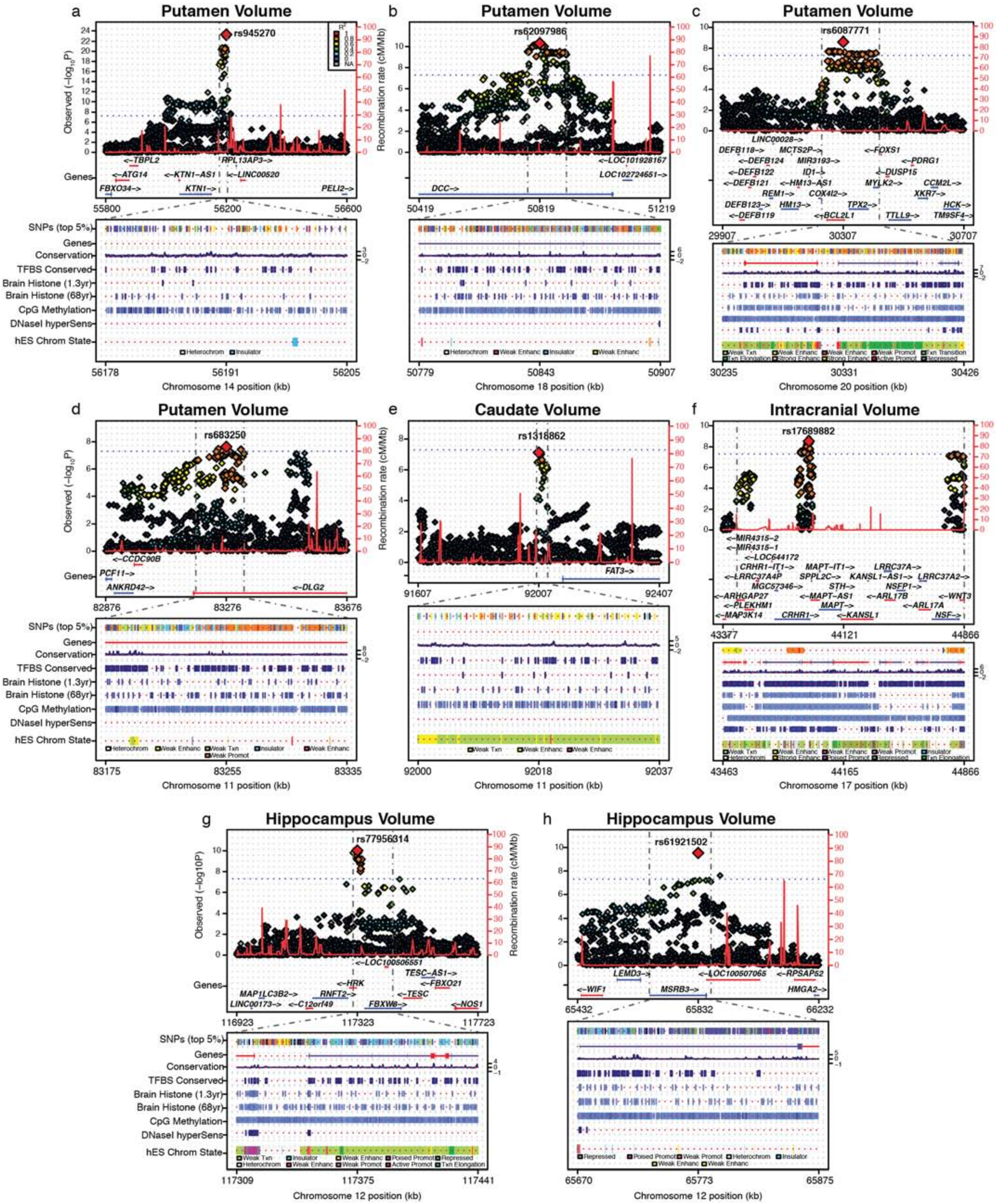
31. Patenaude, B., Smith, S. M., Kennedy, D. N. & Jenkinson, M. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* **56,** 907–922 (2011).
32. Fischl, B. *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33,** 341–355 (2002).
33. Morey, R. A. *et al.* Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. *Hum. Brain Mapp.* **31,** 1751–1762 (2010).
34. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34,** 816–834 (2010).
35. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genet.* **44,** 955–959 (2012).
36. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genet.* **30,** 97–101 (2002).
37. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).
38. Nyholt, D. R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* **74,** 765–769 (2004).
39. Boker, S. *et al.* OpenMx: an open source extended structural equation modeling framework. *Psychometrika* **76,** 306–317 (2011).
40. Walters, R., Bartels, M. & Lubke, G. Estimating variance explained by all variants in meta-analysis with heterogeneity. *Behav. Genet.* **43,** 543 (2013).
41. So, H. C., Li, M. & Sham, P. C. Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. *Genet. Epidemiol.* **35,** 447–456 (2011).
42. Li, M. X., Gui, H. S., Kwan, J. S. & Sham, P. C. GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am. J. Hum. Genet.* **88,** 283–293 (2011).
43. Li, M. X., Kwan, J. S. & Sham, P. C. HYST: a hybrid set-based test for genome-wide association studies, with application to protein-protein interaction-based association analysis. *Am. J. Hum. Genet.* **91,** 478–488 (2012).
44. Ramasamy, A. *et al.* Resolving the polymorphism-in-probe problem is critical for correct interpretation of expression QTL studies. *Nucleic Acids Res.* **41,** e88 (2013).
45. Trabzuni, D. *et al.* Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *J. Neurochem.* **119,** 275–282 (2011).
46. Gutman, B. A. *et al.* Maximizing power to track Alzheimer's disease and MCI progression by LDA-based weighting of longitudinal ventricular surface features. *Neuroimage* **70,** 386–401 (2013).
47. Gutman, B. A., Wang, Y. L., Rajagopalan, P., Toga, A. W. & Thompson, P. M. Shape matching with medial curves and 1-d group-wise registration. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, 716–719 (2012).
48. Schumann, G. *et al.* The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol. Psychiatry* **15,** 1128–1139 (2010).
49. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20,** 110–121 (2010).
50. Cheung, I. *et al.* Developmental regulation and individual differences of neuronal H3K4me3 epigenomes in the prefrontal cortex. *Proc. Natl Acad. Sci. USA* **107,** 8824–8829 (2010).
51. Maunakea, A. K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466,** 253–257 (2010).
52. Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132,** 311–322 (2008).
53. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).
54. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55,** 997–1004 (1999).
55. Hager, R., Lu, L., Rosen, G. D. & Williams, R. W. Genetic architecture supports mosaic brain evolution and independent brain-body size regulation. *Nat. Commun.* **3,** 1079 (2012).
56. Schmucker, D. & Chen, B. Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. *Genes Dev.* **23,** 147–156 (2009).
57. Brunet, A., Datta, S. R. & Greenberg, M. E. Transcription-dependent and -independent control of neuronal survival by the PI3K-Akt signaling pathway. *Curr. Opin. Neurobiol.* **11,** 297–305 (2001).



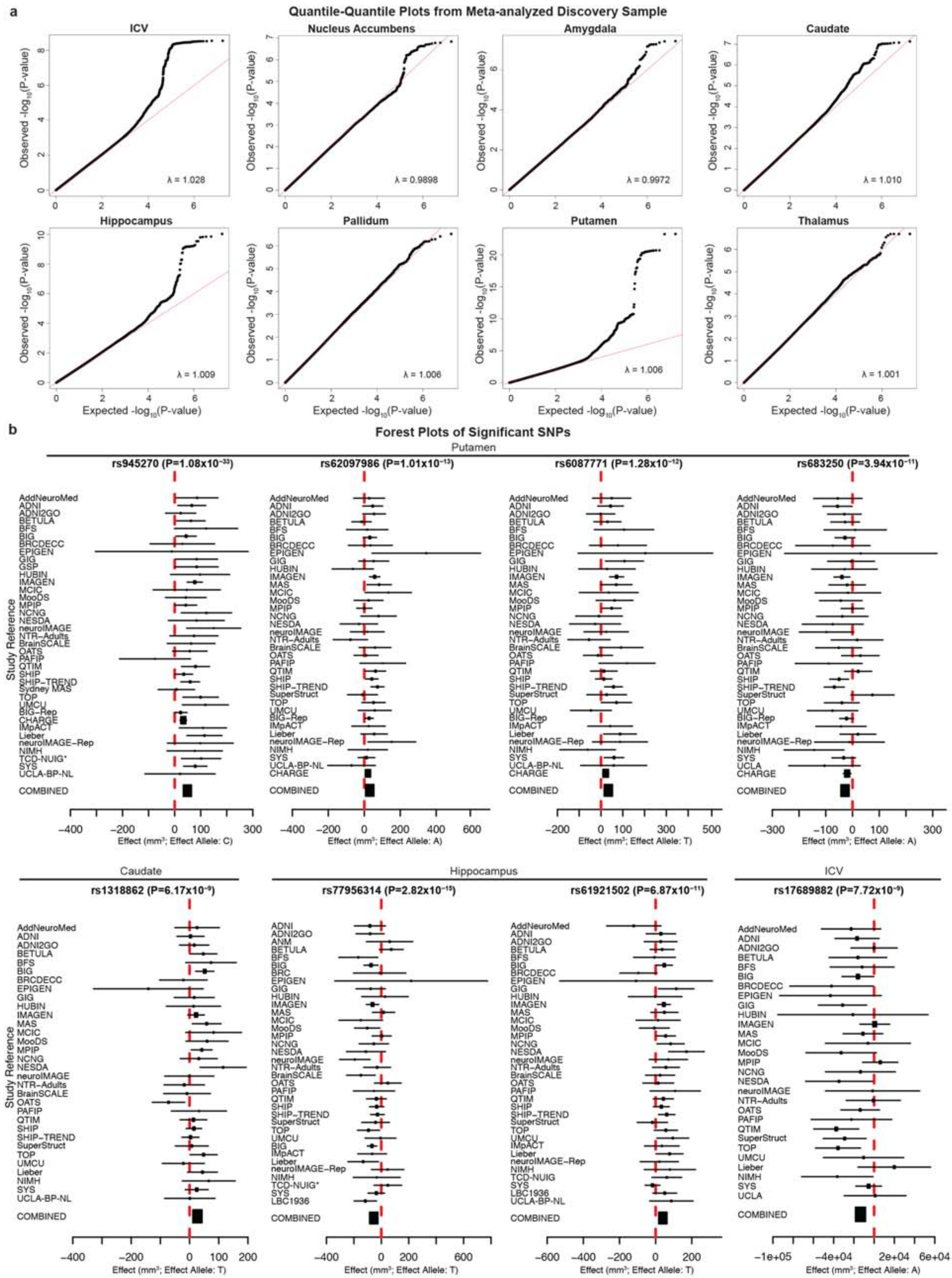
Extended Data Figure 1 | Outline of the genome-wide association meta-analysis. Structural T_1-weighted brain MRI and biological specimens for DNA extraction were acquired from each individual at each site. Imaging protocols were distributed to and completed by each site for standardized automated segmentation of brain structures and calculation of the ICV. Volumetric phenotypes were calculated from the segmentations. Genome-wide genotyping was completed at each site using commercially available chips. Standard imputation protocols to the 1000 Genomes reference panel (phase 1, version 3) were also distributed and completed at each site. Each site completed genome-wide association for each of the eight volumetric brain phenotypes with the listed covariates. Statistical results from GWAS files were uploaded to a central site for quality checking and fixed effects meta-analysis.



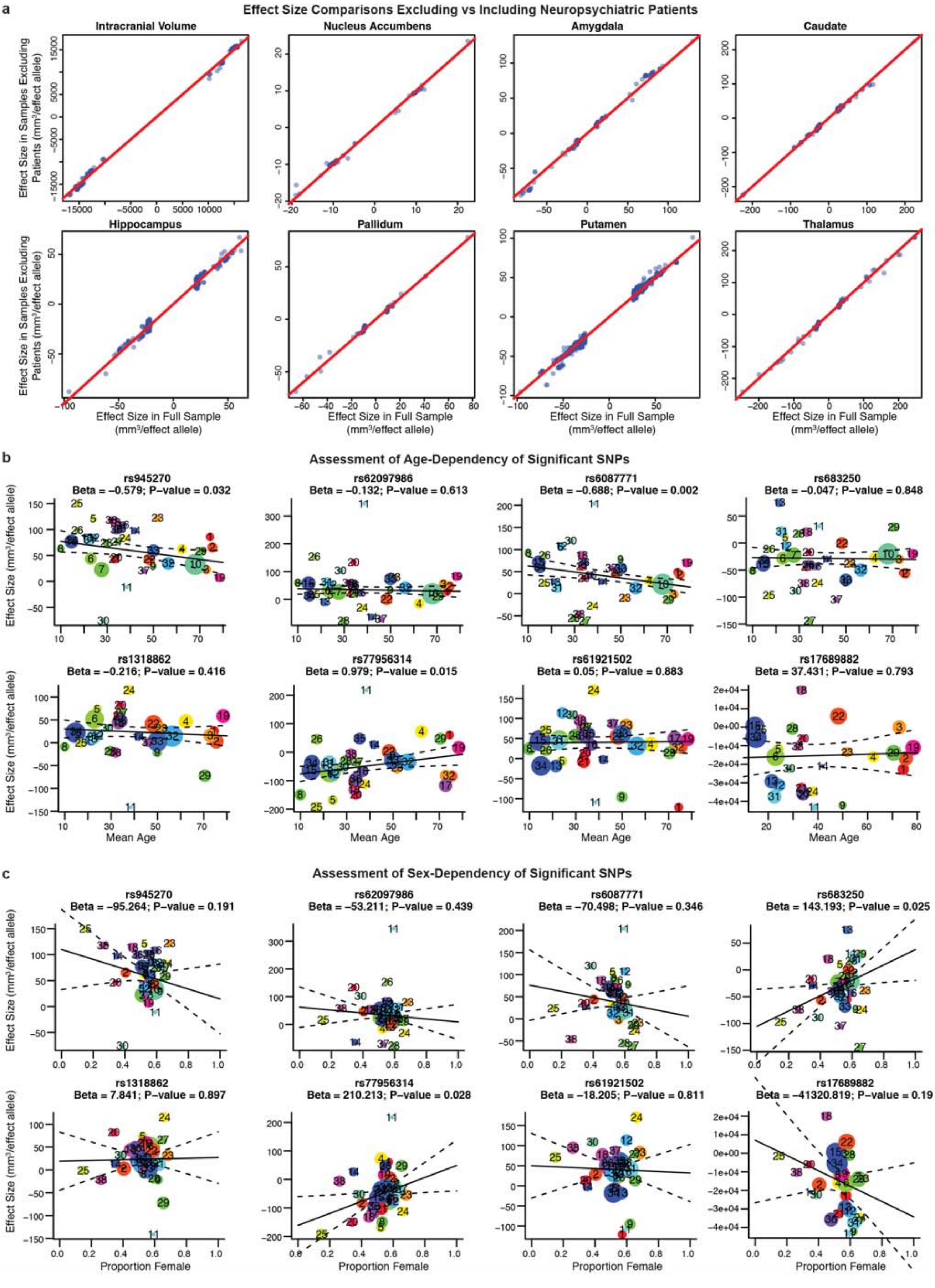
Extended Data Figure 2 | Ancestry inference via multi-dimensional scaling plots. Multi-dimensional scaling (MDS) plots of the discovery cohorts to HapMap III reference panels of known ancestry are displayed. Ancestry is generally homogeneous within each group. In all discovery samples any individuals with non-European ancestry were excluded before association. The axes have been flipped to the same orientation for each sample for ease of comparison. ASW, African ancestry in southwest USA; CEU, Utah residents with northern and western European ancestry from the CEPH collection; CHD, Chinese in metropolitan Denver, Colorado; GIH, Gujarati Indians in Houston, Texas; LWK, Luhya in Webuye, Kenya; MEX, Mexican ancestry in Los Angeles, California; MKK, Maasai in Kinyawa, Kenya; TSI, Tuscans in Italy; YRI, Yoruba in Ibadan, Nigeria.



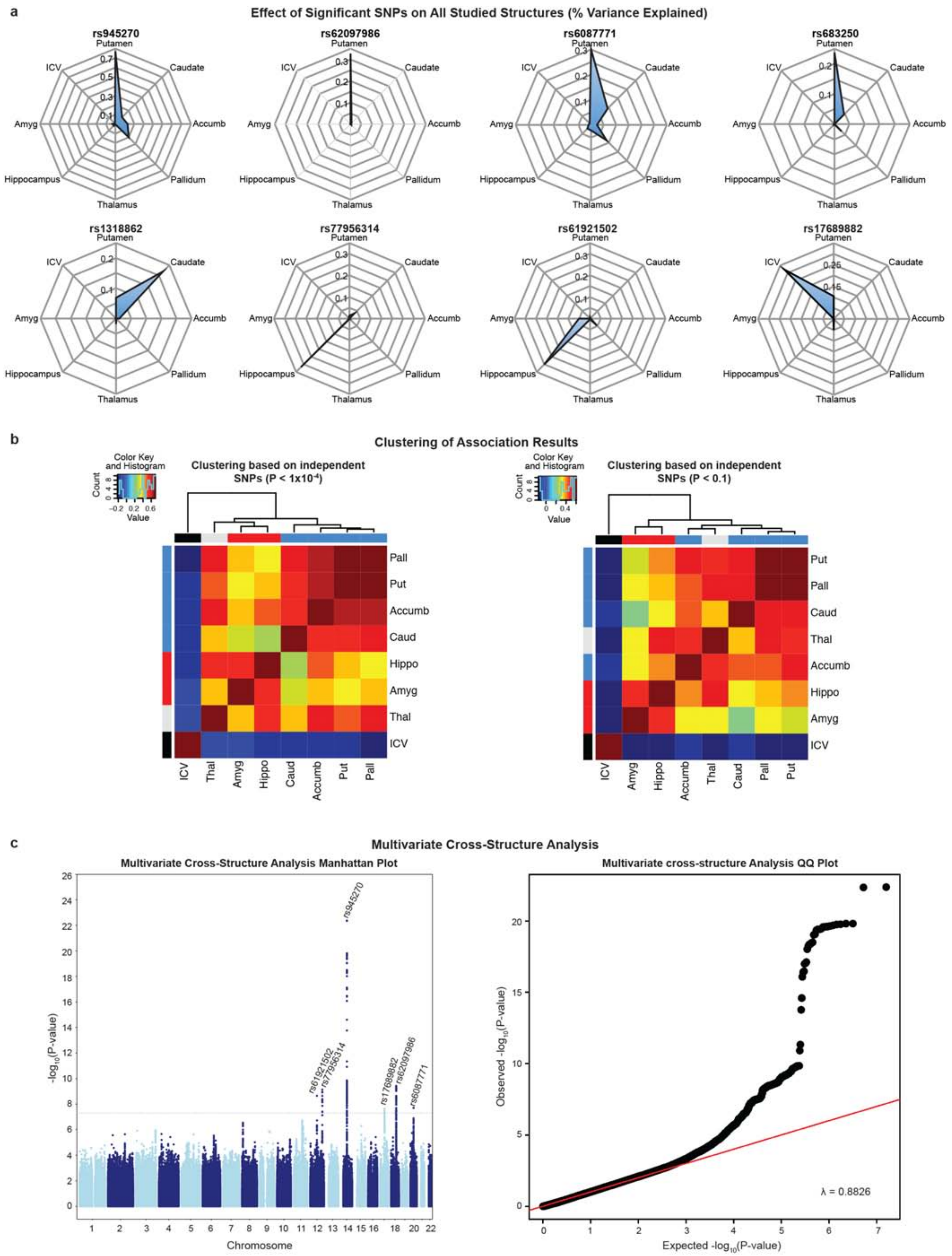
Extended Data Figure 3 | Genomic function is annotated near novel genome-wide significant loci. **a**–**h**, For each panel, zoomed-in Manhattan plots (±400 kb from top SNP) are shown with gene models below (GENCODE version 19). Plots below are zoomed to highlight the genomic region that probably contains the causal variant(s) ($r^2 > 0.8$ from the top SNP). Genomic annotations from the UCSC browser and ENCODE are displayed to indicate potential functionality (see Methods for detailed track information). SNP coverage is low in **f** owing to a common genetic inversion in the region. Each plot was made using the LocusTrack software (http://gump.qimr.edu.au/general/gabrieC/LocusTrack/).



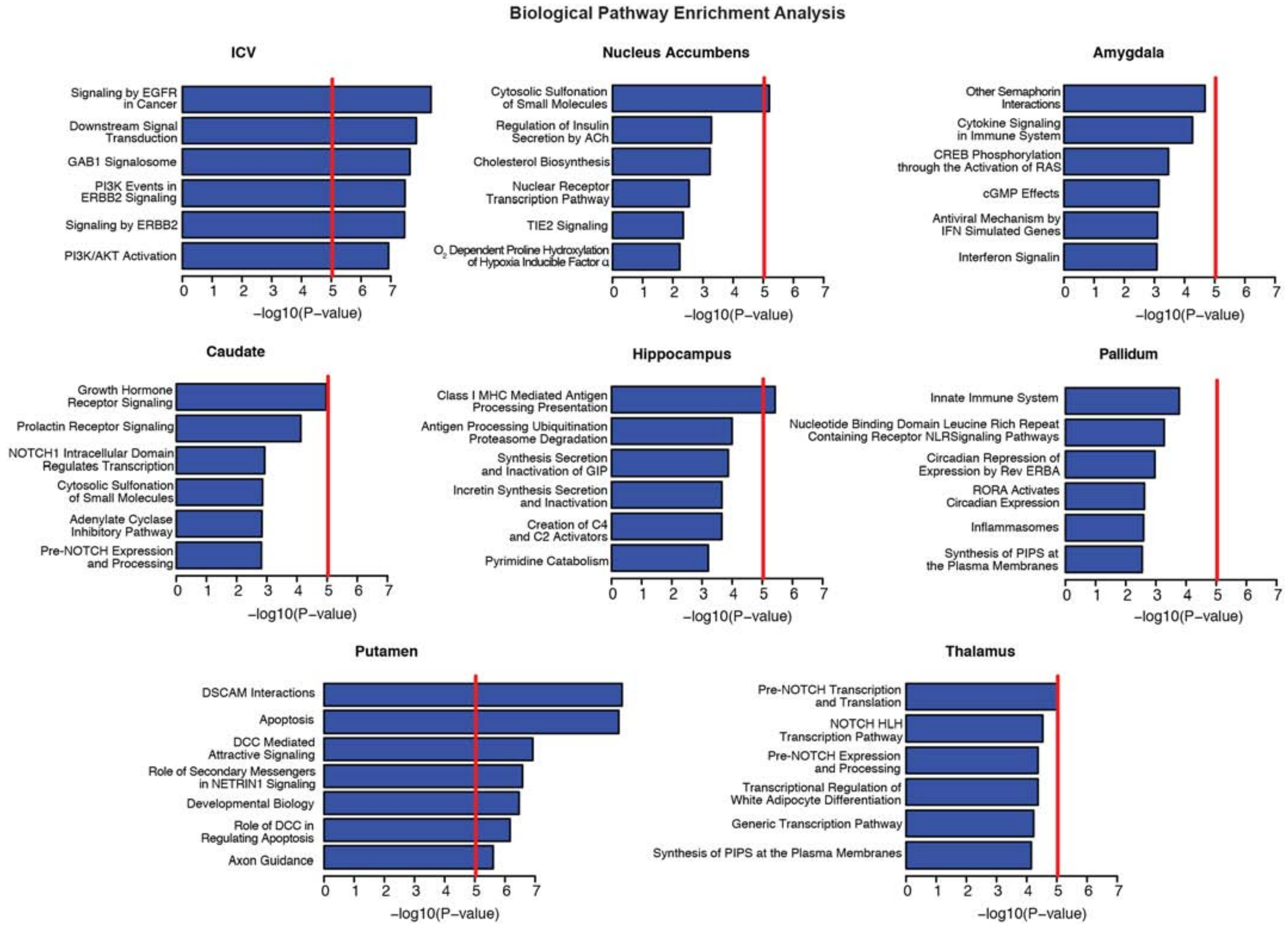
Extended Data Figure 4 | Quantile–quantile and forest plots from meta-analysis of discovery cohorts. **a**, Quantile–quantile plots show that the observed *P* values only deviate from the expected null distribution at the most significant values, indicating that population stratification or cryptic relatedness are not unduly inflating the results. This is quantified through the genomic control parameter (λ; which evaluates whether the median test statistic deviates from expected)[54]. λ values near 1 indicate that the median test statistic is similar to those derived from a null distribution. Corresponding meta-analysis Manhattan plots can be found in Fig. 1. **b**, Forest plots show the effect at each of the contributing sites to the meta-analysis. The size of the dot is proportional to the sample size, the effect is shown by the position on the *x* axis, and the standard error is shown by the line. Sites with an asterisk indicate the genotyping of a proxy SNP (in perfect linkage disequilibrium calculated from 1000 Genomes) for replication.


a
Effect Size Comparisons Excluding vs Including Neuropsychiatric Patients
Intracranial Volume
Nucleus Accumbens
Amygdala
Caudate
Hippocampus
Pallidum
Putamen
Thalamus
Effect Size in Samples Excluding Patients (mm³/effect allele)
Effect Size in Full Sample (mm³/effect allele)
b
Assessment of Age-Dependency of Significant SNPs
rs945270
Beta = −0.579; P−value = 0.032
rs62097986
Beta = −0.132; P−value = 0.613
rs6087771
Beta = −0.688; P−value = 0.002
rs683250
Beta = −0.047; P−value = 0.848
rs1318862
Beta = −0.216; P−value = 0.416
rs77956314
Beta = 0.979; P−value = 0.015
rs61921502
Beta = 0.05; P−value = 0.883
rs17689882
Beta = 37.431; P−value = 0.793
Effect Size (mm³/effect allele)
Mean Age
c
Assessment of Sex-Dependency of Significant SNPs
rs945270
Beta = −95.264; P−value = 0.191
rs62097986
Beta = −53.211; P−value = 0.439
rs6087771
Beta = −70.498; P−value = 0.346
rs683250
Beta = 143.193; P−value = 0.025
rs1318862
Beta = 7.841; P−value = 0.897
rs77956314
Beta = 210.213; P−value = 0.028
rs61921502
Beta = −18.205; P−value = 0.811
rs17689882
Beta = −41320.819; P−value = 0.19
Effect Size (mm³/effect allele)
Proportion Female

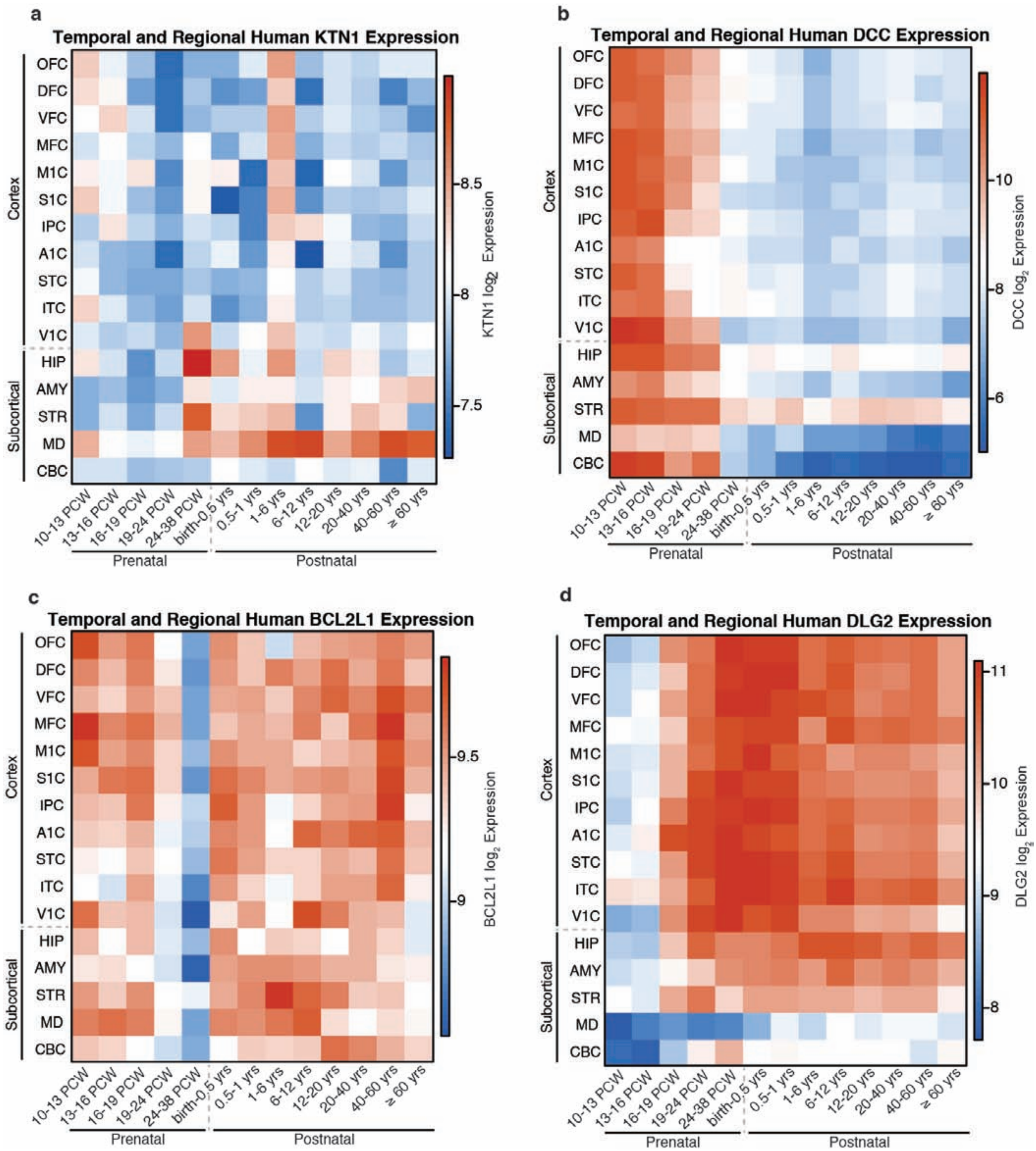
Extended Data Figure 5 | Influence of patients with neuropsychiatric disease, age and gender on association results. **a**, Scatterplot of effect sizes including and excluding patients with neuropsychiatric disorders for nominally significant SNPs. For each of the eight volumetric phenotypes, SNPs with $P < 1 \times 10^{-5}$ in the full discovery set meta-analysis were also evaluated excluding the patients. The beta values from regression, a measure of effect size, are plotted (blue dots) along with a line of equivalence between the two conditions (red line). The correlation between effect sizes with and without patients was very high ($r > 0.99$), showing that the SNPs with significant effects on brain structure are unlikely to be driven by the diseased individuals. **b**, Meta-regression comparison of effect size with mean age at each site. Each site has a corresponding number and coloured dot in each graph. The size of each dot is based on the standard error such that bigger sites with more definitive estimates have larger dots (and more influence on the meta-regression). The age range of participants covered most of the lifespan (9–97 years), but only one of these eight loci showed a significant relationship with the mean age of each cohort (rs608771 affecting putamen volume). **c**, Meta-regression comparison of effect size with the proportion of females at each site. No loci showed evidence of moderation by the proportion of females in a given sample. However, the proportion of females at each site has a very restricted range, so results should be interpreted with caution. Plotted information follows the same convention as described in **b**. The sites are numbered in the following order: (1) AddNeuroMed, (2) ADNI, (3) ADNI2GO, (4) BETULA, (5) BFS, (6) BIG, (7) BIG-Rep, (8) BrainSCALE, (9) BRCDECC, (10) CHARGE, (11) EPIGEN, (12) GIG, (13) GSP, (14) HUBIN, (15) IMAGEN, (16) IMpACT, (17) LBC1936, (18) Lieber, (19) MAS, (20) MCIC, (21) MooDS, (22) MPIP, (23) NCNG, (24) NESDA, (25) neuroIMAGE, (26) neuroIMAGE-Rep, (27) NIMH, (28) NTR-Adults, (29) OATS, (30) PAFIP, (31) QTIM, (32) SHIP, (33) SHIP-TREND, (34) SYS, (35) TCD-NUIG, (36) TOP, (37) UCLA-BP-NL and (38) UMCU.


a
Effect of Significant SNPs on All Studied Structures (% Variance Explained)
rs945270
rs62097986
rs6087771
rs683250
rs1318862
rs77956314
rs61921502
rs17689882
Putamen
Caudate
Accumb
Pallidum
Thalamus
Hippocampus
Amyg
ICV
b
Clustering of Association Results
Color Key and Histogram
Count
Value
Clustering based on independent SNPs (P < 1x10⁻⁴)
Clustering based on independent SNPs (P < 0.1)
Pall
Put
Accumb
Caud
Hippo
Amyg
Thal
ICV
c
Multivariate Cross-Structure Analysis
Multivariate Cross-Structure Analysis Manhattan Plot
-log10(P-value)
Chromosome
rs945270
rs61921502
rs77956314
rs17689882
rs62097986
rs6087771
Multivariate cross-structure Analysis QQ Plot
Observed -log10(P-value)
Expected -log10(P-value)
λ = 0.8826

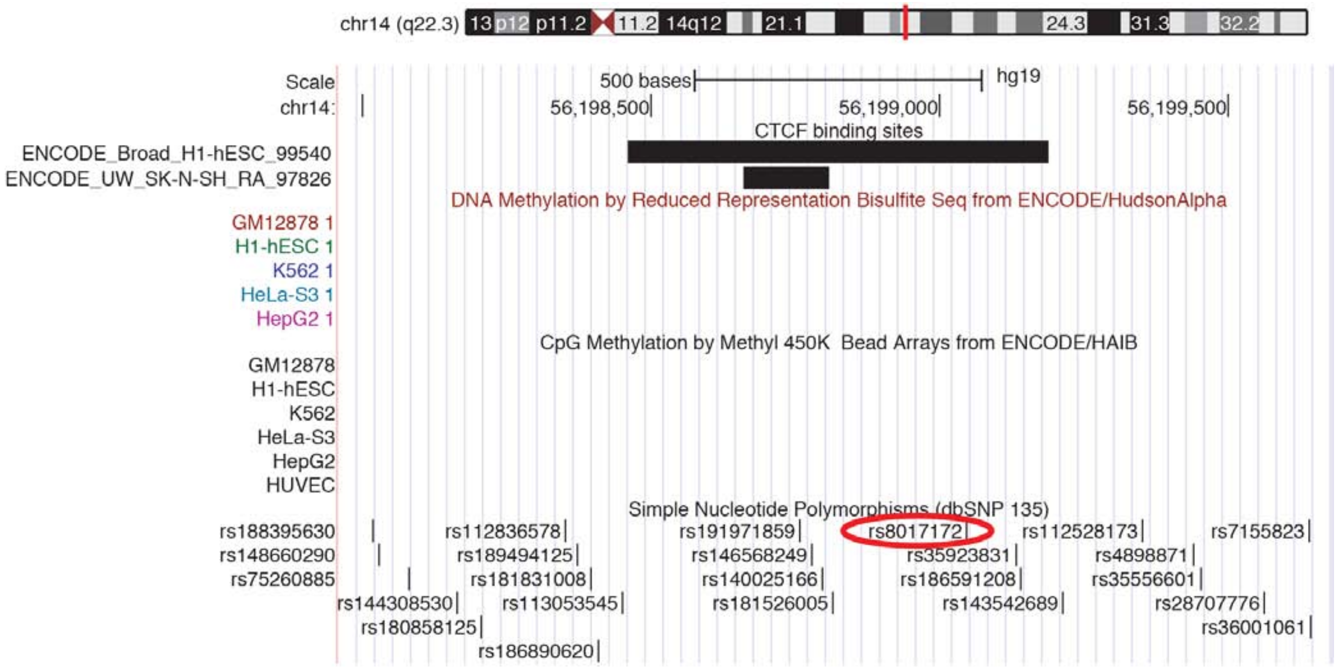
Extended Data Figure 6 | Cross-structure analyses. **a**, Radial plots of effect sizes from the discovery sample for all genome-wide significant SNPs identified in this study. Plots indicate the effect of each genetic variant, quantified as percentage variance explained, on the eight volumetric phenotypes studied. As expected, the SNPs identified with influence on a phenotype show the highest effect size for that phenotype: putamen volume (rs945270, rs62097986, rs608771 and rs683250), hippocampal volume (rs77956314 and rs61921502), caudate volume (rs1318862) and ICV (rs17689882). In general much smaller effects are observed on other structures. **b**, Correlation heat map of GWAS test statistics (*t*-values) and hierarchical clustering[55]. Independent SNPs were chosen within an linkage disequilibrium block based on the highest association in the multivariate cross-structure analysis described in Extended Data Fig. 6c. Two heat maps are shown taking only independent SNPs with either $P < 1 \times 10^{-4}$ (left) or $P < 0.01$ (right) in the multivariate cross-structure analysis. Different structures are labelled in developmentally similar regions by the colour bar on the top and side of the heat map including basal ganglia (putamen, pallidum, caudate and accumbens; blue), amygdalo–hippocampal complex (hippocampus and amygdala; red), thalamus (turquoise) and ICV (black). Hierarchical clustering showed that developmentally similar regions have mostly similar genetic influences across the entire genome. The low correlation with the ICV is owing to it being used as a covariate in the subcortical structure GWAS associations. **c**, A multivariate cross-structure analysis of all volumetric brain traits. A Manhattan plot (left) and corresponding quantile–quantile plot (right) of multivariate GWAS analysis of all traits (volumes of the accumbens, amygdala, caudate, hippocampus, pallidum, putamen, thalamus, and ICV) in the discovery data set using the TATES method[9] is shown. Multivariate cross-structure analysis confirmed the univariate analyses (see Table 1), but did not reveal any additional loci achieving cross-structure levels of significance.



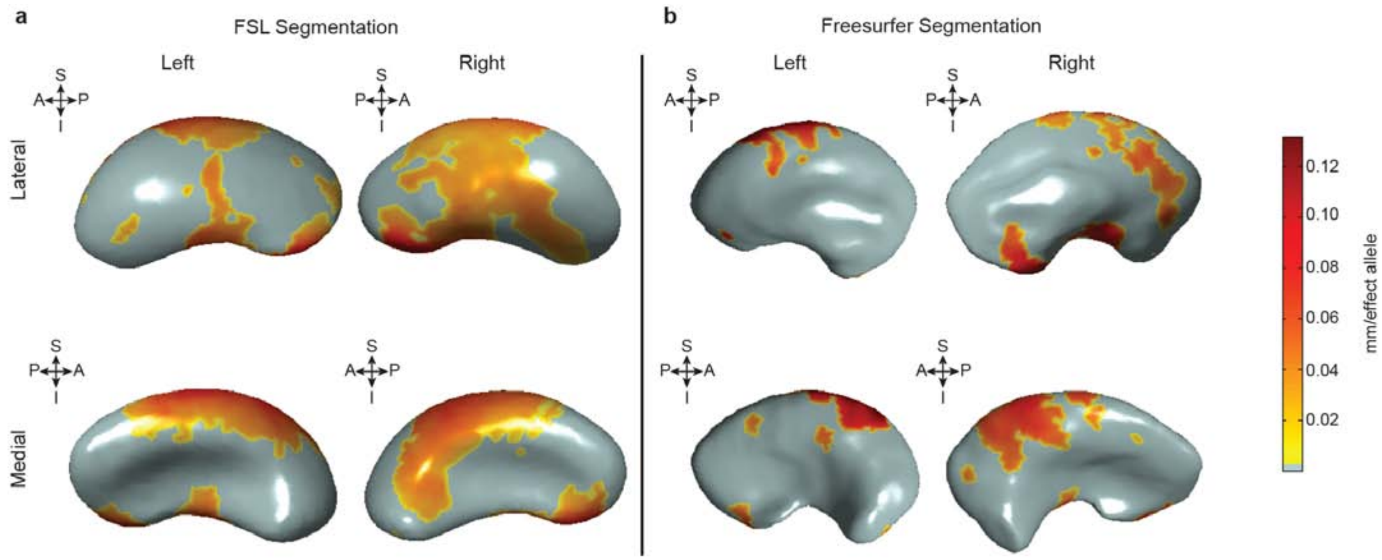
Extended Data Figure 7 | Pathway analysis of GWAS results for each brain structure. A pathway analysis was performed on each brain volume GWAS using KGG[42] to conduct gene-based tests and the Reactome database for pathway definition[43]. Pathway-wide significance was calculated using a Bonferroni correction threshold accounting for the number of pathways and traits tested such that $P_{\mathrm{thresh}} = 0.05/(671 \text{ pathways} \times 7 \text{ independent traits}) = 1.06 \times 10^{-5}$ and is shown here as a red line. The number of independent traits was calculated by accounting for the non-independence of each of the eight traits examined (described in the Methods). Variants that influence the putamen were clustered near genes known to be involved in DSCAM interactions, neuronal arborization and axon guidance[56]. Variants that influence intracranial volume are clustered near genes involved in EGFR and phosphatidylinositol-3-OH kinase (PI(3)K)/AKT signalling pathways, known to be involved in neuronal survival[57]. All of these represent potential mechanisms by which genetic variants influence brain structure. It is important to note that the hybrid set-based test (HYST) method for pathway analysis used here can be strongly influenced by a few highly significant genes, as was the case for putamen hits in which *DCC* and *BCL2L1* were driving the pathway results.



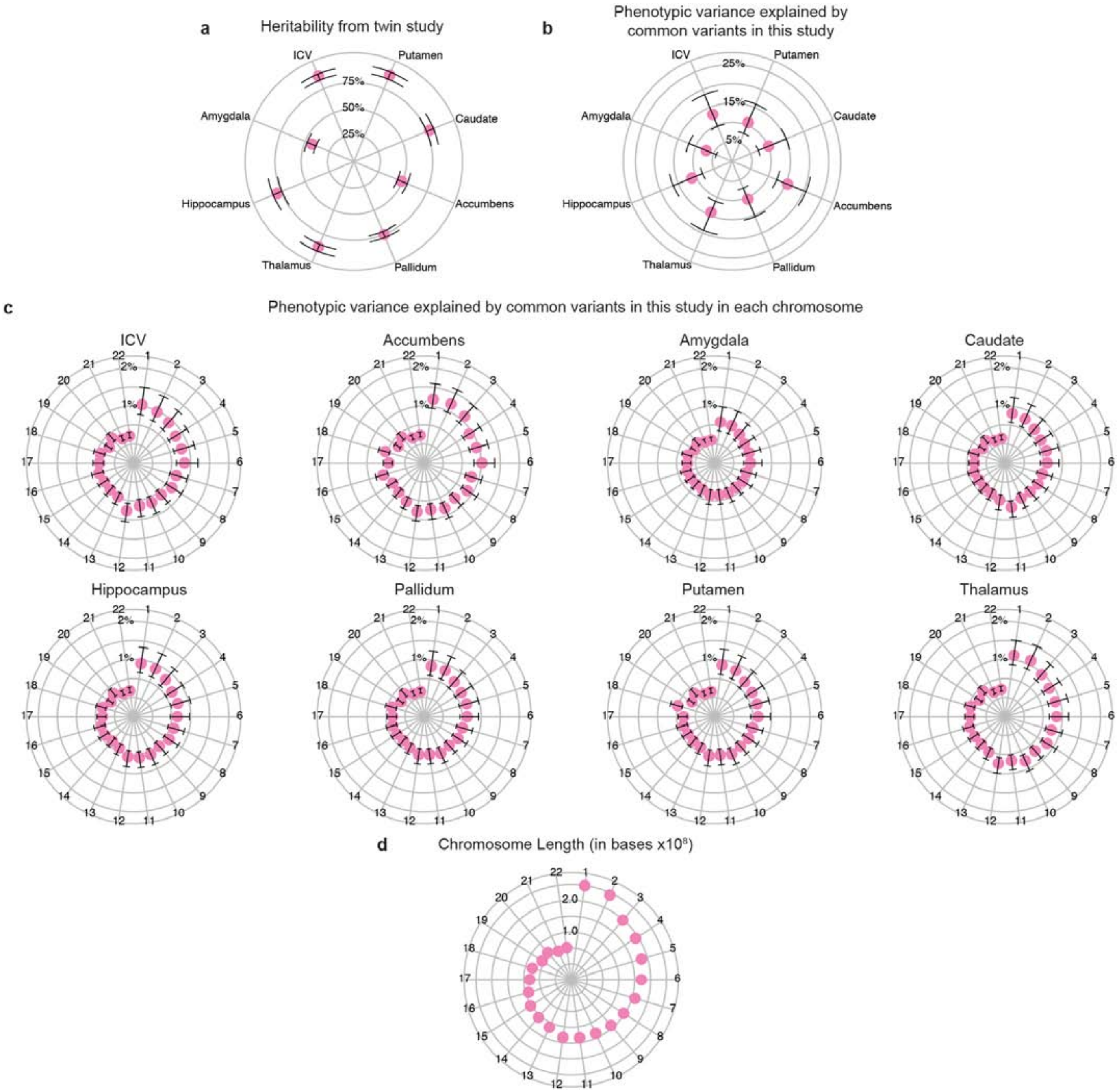
Extended Data Figure 8 | Spatio-temporal maps showing expression of genes near the four significant putamen loci over time and throughout regions of the brain. Spatio-temporal gene expression[13] was plotted as normalized $\log_2$ expression. Different areas of the neocortex (A1C, primary auditory cortex; DFC, dorsolateral prefrontal cortex; IPC, posterior inferior parietal cortex; ITC, inferior temporal cortex; MFC, medial prefrontal cortex; M1C, primary motor cortex; OFC, orbital prefrontal cortex; STC, superior temporal cortex; S1C, primary somatosensory cortex; VFC, ventrolateral prefrontal cortex; V1C, primary visual cortex) as well as subcortical areas (AMY, amygdala; CBC, cerebellar cortex; HIP, hippocampus; MD, mediodorsal nucleus of the thalamus; STR, striatum) are plotted from 10 post-conception weeks (PCW) to more than 60 years old. Genes that probably influence putamen volume are expressed in the striatum at some point during the lifespan. After late fetal development, *KTN1* is expressed in the human thalamus, striatum and hippocampus and is more highly expressed in the striatum than the cortex. Most genes seem to have strong gradients of expression across time, with *DCC* most highly expressed during early prenatal life, and *DLG2* most highly expressed at mid-fetal periods and throughout adulthood. *BCL2L1*, which inhibits programmed cell death, has decreased striatal expression at the end of neurogenesis (24–38 PCW), a period marked by increased apoptosis in the putamen[15].



Extended Data Figure 9 | CTCF-binding sites in the vicinity of the putamen locus marked by rs945270. CTCF-binding sites from the ENCODE project are displayed from the database CTCFBSDB 2.0 (ref. 23) from two different cell types: embryonic stem cells (track ENCODE_Broad_H1-hESC_99540) and a neuroblastoma cell line differentiated with retinoic acid (ENCODE_UW_SK-N-SH_RA_97826). A proxy SNP to the top hit within the locus, rs8017172 ($r^2 = 1.0$ to rs945270), lies within a CTCF-binding site called based on ChIP-seq data in the embryonic stem cells and near the binding site in neural SK-N-SH cells. As this is the lone chromatin mark in the intergenic region (see Extended Data Fig. 3), it suggests that the variant may disrupt a CTCF-binding site and thereby influence transcription of surrounding genes.



Extended Data Figure 10 | Shape analysis in 1,541 young healthy subjects shows consistent deformations of the putamen regardless of segmentation protocol. **a**, **b**, The distance from a medial core to surfaces derived from FSL FIRST (**a**; identical to Fig. 2c) or FreeSurfer (**b**) segmentations was derived in the same 1,541 subjects. Each copy of the rs945270-C allele was significantly associated with an increased width in coloured areas (false discovery rate corrected at $q = 0.05$) and the degree of deformation is labelled by colour. The orientation is indicated by arrows. A, anterior; I, inferior; P, posterior; S, superior. Shape analysis in both software suites gives statistically significant associations in the same direction. Although the effects are more widespread in the FSL segmentations, FreeSurfer segmentations also show overlapping regions of effect, which appears strongest in anterior and superior sections.



Extended Data Figure 11 | The phenotypic variance explained by all common variants in this study. **a**, Twin-based heritability (with 95% confidence intervals), measuring additive genetic influences from both common and rare variation, is shown for comparison with common variant based heritability (see Methods). **b**, The median estimated percentage of phenotypic variance explained by all SNPs (and 95% confidence interval) is given for each brain structure studied[41]. The full genome-wide association results from common variants explain approximately 7–15% of variance depending on the phenotype. **c**, The median estimated variance explained by each chromosome is shown for each phenotype. **d**, Some chromosomes explain more variance than would be expected by their length, for example chromosome 18 in the case of the putamen, which contains the *DCC* gene.

doi:10.1038/nature14361

The maternal-age-associated risk of congenital heart disease is modifiable

Claire E. Schulkey[1], Suk D. Regmi[1], Rachel A. Magnan[1], Megan T. Danzo[1], Herman Luther[1], Alayna K. Hutchinson[1], Adam A. Panzer[1], Mary M. Grady[1], David B. Wilson[1,2] & Patrick Y. Jay[1,3]

Maternal age is a risk factor for congenital heart disease even in the absence of any chromosomal abnormality in the newborn[1–7]. Whether the basis of this risk resides with the mother or oocyte is unknown. The impact of maternal age on congenital heart disease can be modelled in mouse pups that harbour a mutation of the cardiac transcription factor gene *Nkx2-5* (ref. 8). Here, reciprocal ovarian transplants between young and old mothers establish a maternal basis for the age-associated risk in mice. A high-fat diet does not accelerate the effect of maternal ageing, so hyperglycaemia and obesity do not simply explain the mechanism. The age-associated risk varies with the mother's strain background, making it a quantitative genetic trait. Most remarkably, voluntary exercise, whether begun by mothers at a young age or later in life, can mitigate the risk when they are older. Thus, even when the offspring carry a causal mutation, an intervention aimed at the mother can meaningfully reduce their risk of congenital heart disease.

Congenital heart disease remains a leading cause of childhood morbidity and mortality despite dramatic clinical advances. Discoveries regarding the pathogenic mechanisms have likewise abounded, but translating knowledge about the causes of the disease in the embryo to improve outcomes will not be simple. The ideal would be to develop a broadly implemented prevention strategy, just as folic acid is prescribed to expectant mothers to prevent neural tube defects[9]. Therefore, we have focused upon genetic or environmental modifiers of heart defects caused by *Nkx2-5* haploinsufficiency in the mouse embryo[8,10]. Mutations of the cardiac transcription factor NKX2-5 that cause congenital heart disease were first discovered in humans[11,12]. The identification of modifiers, which affect risk but do not cause disease *per se*, can point the way towards therapies that do not necessarily target the main cause[13] but are effective nonetheless.

A substantial fraction of the ventricular septal defects (VSDs) seen among *Nkx2-5*$^{+/-}$ mice can be attributed to the effect of modifiers[8]. These modifiers include genetic polymorphisms that quantitatively affect penetrance[8,10]. In the C57BL/6N × FVB/N hybrid strain background, maternal age also affects the risk of VSD in *Nkx2-5*$^{+/-}$ but not wild-type pups; Extended Data Fig. 1 depicts the breeding scheme. The risk is independent of genetic polymorphisms in the offspring and not due to chromosomal aneuploidy[8]. These laboratory and epidemiological observations motivated investigation of the maternal age effect in the mouse model.

The basis of the maternal age effect could reside in either the mother or oocyte. To determine which, we performed reciprocal ovarian transplants between young and old mothers that were first generation (F1) hybrids of the inbred strains C57BL/6N and FVB/N (Fig. 1). F1 parents were bred to produce F2 offspring. The incidence of VSD was significantly greater among the *Nkx2-5*$^{+/-}$ offspring of older mothers bearing young ovaries. Next, we calculated the number of VSDs that would be expected for either a maternal or oocyte basis by looking at the effect of maternal age that was quantified in the 2,262 *Nkx2-5*$^{+/-}$ offspring of C57BL/6N × FVB/N F1 mothers who ovulated from their native ovaries. The observed numbers in each transplanted group were consistent with a maternal basis of the age effect (Fig. 1). We focused on VSDs, the most common defect in *Nkx2-5*$^{+/-}$ mice, because their frequency provides sufficient power for the statistical analyses we carried out[8,10]. Atrial septal defects (ASDs), the second most common defect, showed a pattern consistent with a maternal basis, but their numbers were insufficient to draw statistical conclusions (Extended Data Fig. 2). The results point to a maternal pathway that either produces a factor or mediates a process that interacts with cardiac development in the mutant embryo. A harmful factor would rise with age, whereas a protective one would fall.

Factors related to diabetes or obesity are plausible suspects; both conditions are associated with ageing and human congenital heart disease[14].

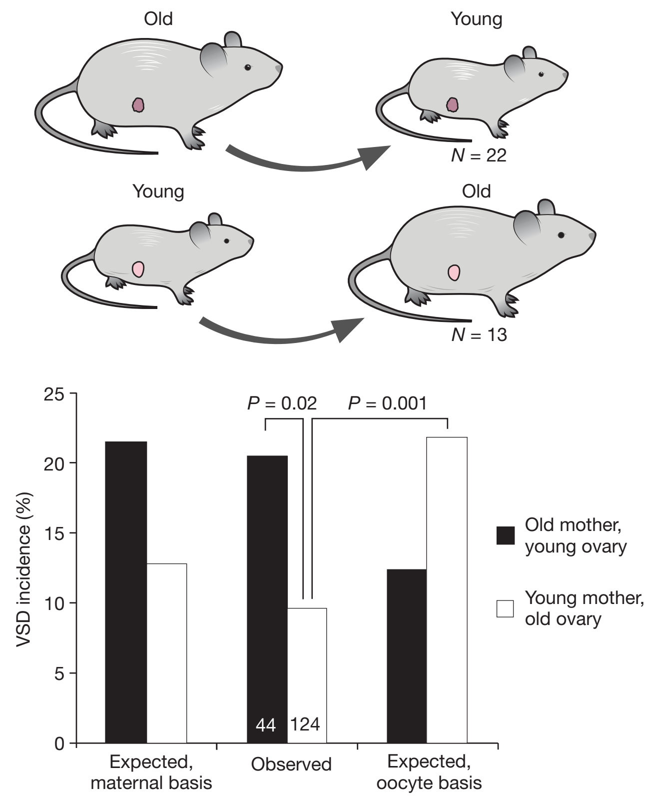


Figure 1 | Reciprocal ovarian transplants between young and old mothers localize the basis of the maternal-age-associated risk to the mother. The incidence of VSD for the offspring of old mothers with young ovaries is significantly greater than that of young mothers with old ovaries. The observed incidence in the offspring of recipient mothers matches that expected for a maternal but not an oocyte basis of the age effect. The observed and expected incidences were compared in a chi-squared goodness-of-fit test. The number of recipient mothers and the number of pups in each age group are shown.

[1]Department of Pediatrics, Washington University School of Medicine, St Louis, Missouri 63110 USA. [2]Department of Developmental Biology, Washington University School of Medicine, St Louis, Missouri 63110 USA. [3]Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63110 USA.

To assess their contribution to the maternal-age-associated risk, we placed young females on a high-fat diet 4 weeks before mating. The mothers remained on the diet thereafter. Young mothers on a high-fat diet had impaired glucose tolerance and fasting hyperglycaemia, yet their offspring had an incidence of VSD equal to those of mothers on a normal, chow diet (Fig. 2a, b, d). In addition, old mothers developed marked adiposity. Still, the incidence of VSD in their offspring was not significantly increased (Fig. 2c, d, and Extended Data Fig. 3). Despite large effects on glucose homeostasis and obesity, a high-fat diet did not significantly increase the maternal-age-associated risk of congenital heart defects (Fig. 2e and Extended Data Fig. 4). A logistic regression analysis that considered an interaction between diet and age also failed to detect an effect. Any potential effect of the high-fat diet on the maternal-age-associated risk is small relative to the effects on glycaemic indices or body composition.

Most epidemiological studies identify maternal age as a risk factor for congenital heart disease, but the risk is not always detected, perhaps because of genetic variation among populations[6,15,16]. We thus compared the maternal-age-associated risk in three mouse strain backgrounds: the inbred C57BL/6N strain, and the F1 hybrids of C57BL/6N × A/J and C57BL/6N × FVB/N mice (Extended Data Fig. 1). The inbred C57BL/6N strain bears the greatest risk. That risk is significantly greater than in the C57BL/6N × A/J hybrid, which shows no significant risk. The maternal-age-associated risk in the C57BL/6N × FVB/N hybrid is significant and intermediate to the other two backgrounds (Fig. 3). The maternal-age-associated risk is a quantitative genetic trait. For example, the net effect of A/J polymorphisms is to reduce the risk associated with ageing, whereas C57BL/6N polymorphisms generally increase risk. Genetic polymorphisms may affect the activity of the maternal factor hypothesized to interact with embryonic cardiac development.

A/J and C57BL/6 mice are known for their phenotypic differences in a number of complex metabolic traits, including the effects of diet[17,18]. We thus wondered whether exercise, which has beneficial effects on metabolism, could decrease the risk of congenital heart disease, just as A/J polymorphisms decrease the risk of adverse metabolic phenotypes. Running wheels were placed in the cages of C57BL/6N × FVB/N F1 mice at the onset of breeding, when the females were 4 weeks old. The mice were allowed to run *ad libitum* during their entire reproductive

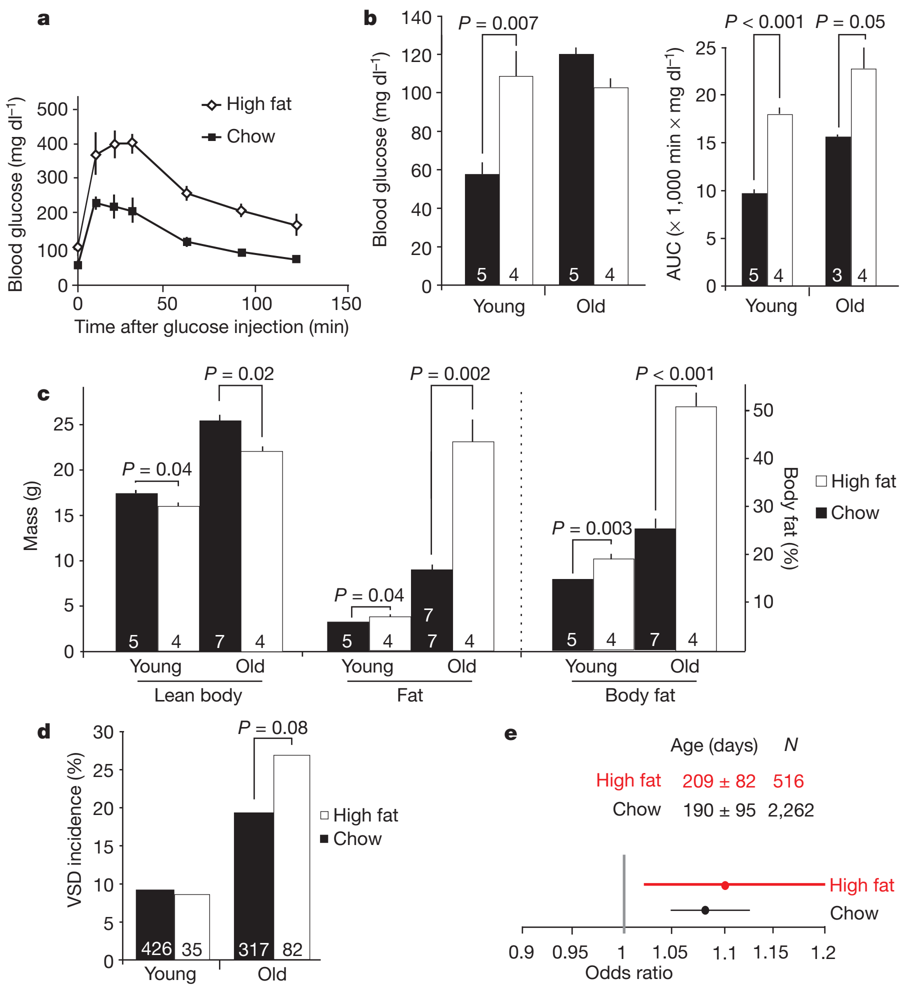


Figure 2 | A high-fat diet does not accelerate the onset of the maternal age effect. **a**, A high-fat diet causes a higher peak glucose and impaired clearance in glucose tolerance tests. Data are shown here from young mothers. **b**, A high-fat diet induces hyperglycaemia in young mothers and impaired glucose tolerance in young and old mothers, as quantified by the area under the curve (AUC) in glucose tolerance tests. **c**, Lean body mass decreases with a high-fat diet, while fat mass and percentage increases. Marked adiposity develops in old mothers on a high-fat diet. **d**, The incidences of VSD in the offspring of young mothers are the same whether on a high-fat or normal diet. The incidence for the offspring of old mothers on a high-fat diet is not significantly increased. Young and old mothers are defined as <100 and >300 days old, respectively. Glucose tolerance tests and magnetic resonance imaging (MRI) quantification of body composition were performed on a subset of the 19 high-fat- or 156 chow-diet-fed mothers who produced the offspring for analysis. **e**, The maternal-age-associated risk is unaffected by a high-fat diet. Odds ratios are presented with the 95% confidence interval; the relative widths of the confidence intervals are related to the sample sizes. Glucose and body composition data are given as mean ± standard error of the mean (s.e.m.); values were compared in two-sided *t*-tests. The age of the mothers (mean ± standard deviation (s.d.)) and the number of offspring in each cohort are shown.

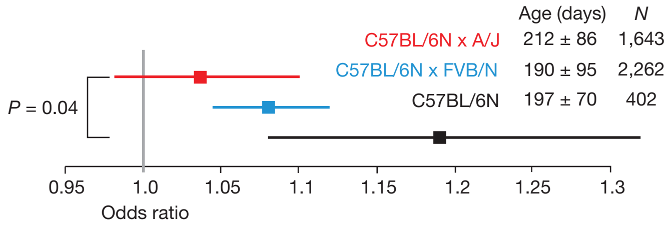


Figure 3 | The maternal-age-associated risk of VSD is a quantitative genetic trait. Risk is significantly greater in the inbred C57BL/6N strain than in the C57BL/6N × A/J F1 hybrid. The C57BL/6N × FVB/N F1 hybrid bears intermediate risk. Odds ratios are presented with the 95% confidence interval; the relative widths of the confidence intervals are related to the sample sizes. The age of mothers (mean ± s.d.) and the number of offspring in each cohort are shown. Odds ratios were compared in a logistic regression analysis that included genetic background as a covariate in addition to maternal age.

lifespan[19]. Early onset exercise did not lower the incidence of VSD in the *Nkx2-5*$^{+/-}$ offspring of young mothers but did in those of old mothers (Fig. 4a). When all the offspring of mothers from young to old age were included in a logistic regression analysis, the maternal-age-associated risk was insignificant in the setting of early onset exercise (Fig. 4b).

Early onset exercise clearly reduces risk for *Nkx2-5*$^{+/-}$ offspring but may be difficult to implement in clinical practice. We therefore assessed the effect of exercise beginning in late adulthood. Running wheels were placed in cages when the mothers were 8 months old. The subsequent incidence of VSD was significantly reduced in the *Nkx2-5*$^{+/-}$ offspring of mothers older than 300 days. Late onset exercise reduced the incidence nearly as much as early onset exercise (Fig. 4a). Consistent with the multifactorial basis of a congenital heart defect, exercise mitigates the fraction of risk associated with maternal ageing but does not eliminate risk entirely, as shown by the equivalent incidences of VSD in the offspring of exercised and sedentary young mothers.

To estimate the duration of exercise sufficient for an effect, the observed incidence of VSD was compared to the expected incidence in a range of intervals binned according to the number of days a mother had exercised by a pup's birth date. Given a 20-day gestational period, 3 months of exercise before conception exerts a detectable effect on incidence (Fig. 4c). Variables not examined here, such as the mother's age at the onset of exercise and the intensity of exercise, could affect the magnitude of the effect. Nevertheless, a modest duration of exercise appears sufficient to reduce the risk of congenital heart disease.

The benefit of exercise for the offspring is not associated with overt or large changes in the mother. Neither early nor late onset exercise significantly affected glucose metabolism, lean or fat body mass or weight in old mothers (Fig. 4d, e and Extended Data Fig. 3). Furthermore, body weight was not independently associated with risk under any of the experimental conditions, that is, sedentary control, high-fat diet, early or late onset exercise. In humans, good evidence indicates that maternal diabetes and obesity are risk factors for congenital heart disease[14]. The fraction of children of mothers who have diabetes or obesity who actually have a heart defect is still small, however, so the true risk factor may be present in just a subset of women. Exercise may ameliorate a risk factor associated with ageing, dysglycaemia or obesity without a noticeable effect on glucose metabolism or body composition in the mouse mother.

The present results reveal a maternal pathway that could potentially be targeted to prevent congenital heart disease in offspring who carry a deleterious mutation. Whether this pathway affects the development of defects other than VSD or that are not caused by *Nkx2-5* mutation remains unanswered, although consistent evidence supports a maternal-age-associated risk for ASD in the mouse (Extended Data Figs 2 and 4) and for other defects in humans. Very large studies are necessary to prove or exclude confidently a maternal-age-associated risk for less common defects. Experimental dissection of the pathway may be a more efficient means forward. The results could indicate whether distinct

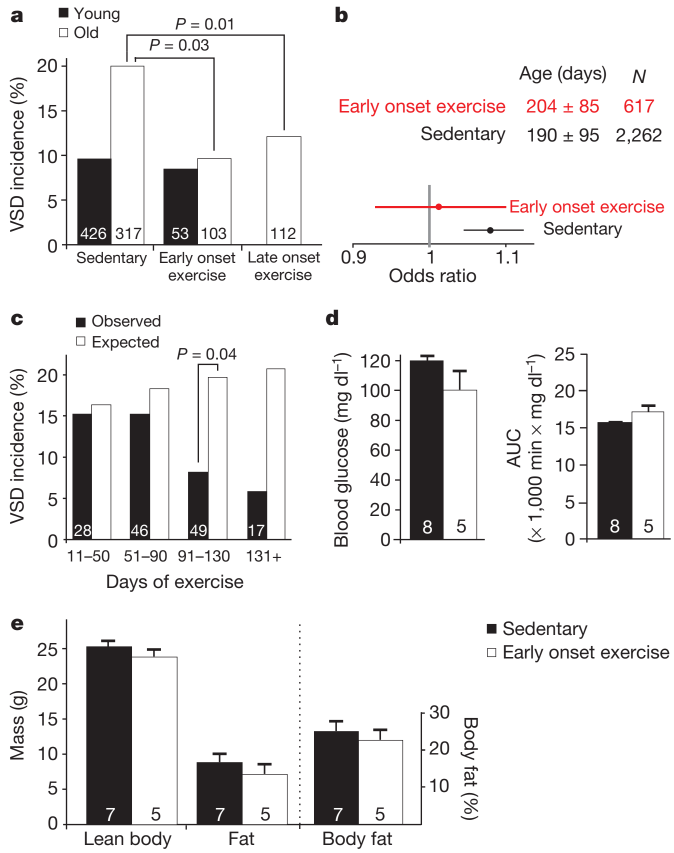


Figure 4 | Exercise can mitigate the risk associated with maternal ageing. **a**–**e**, Running wheels were placed in breeding cages when mothers were 4 weeks or 8 months old (early or late onset exercise groups). **a**, Both early and late-onset exercise decrease the incidence of VSD for the offspring of old compared to those of young mothers. The number of mouse pups in each group is shown. **b**, Early onset exercise makes the risk associated with ageing insignificant. Odds ratios are presented with the 95% confidence interval; the relative widths of the confidence intervals are related to the sample sizes. The age of the mothers (mean ± s.d.) and the number of offspring in each cohort are shown. **c**, The incidence of VSD is shown binned according to the number of days a mother in the late onset group had exercised by the pup's birthdate. The expected incidences are calculated for age-matched, sedentary control mothers. Three months of exercise before conception results in a detectable reduction in the incidence of VSD. **d**, Early onset exercise does not alter fasting glucose levels or glucose tolerance in old mothers. **e**, Lean and fat body mass and body fat percentage are also not appreciably affected. Glucose tolerance tests and MRI quantification of body composition were performed on a subset of the mothers in the early onset exercise and sedentary groups. The data are reported as mean ± s.e.m. The incidences of defects were compared by chi-squared tests.

defects should be analysed separately, and help to focus the aims of human studies. Well-known pathways may not be involved, given that deranged glucose homeostasis and obesity do not exacerbate the maternal age effect. Interestingly, maternal genetic polymorphisms in other metabolic pathways have recently been associated with the risk of conotruncal heart defects in humans[20]. A maternal genetic basis for the age-associated risk in *Nkx2-5*$^{+/-}$ pups likewise supports the concept of gene × gene interactions between the mother and embryo. The present results do not exclude the possibility of other complex molecular, genetic and epigenetic mechanisms, but mapping the genes that affect risk could provide insight into the maternal pathway. In addition, the effect of exercise provides a functional clue. Like ageing, exercise has complex effects on physiology and metabolism. Exercise and aerobic fitness affect the concentration of dozens of metabolites[21–26]. Any of them could be the hypothesized age-related factor. The pertinent factor

or pathway in the mouse, however, must fit a profile delimited by ageing and genetic background.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 August 2014; accepted 27 February 2015.

Published online 1 April 2015.

1. Forrester, M. B. & Merz, R. D. Descriptive epidemiology of selected congenital heart defects, Hawaii, 1986–1999. *Paediatr. Perinat. Epidemiol.* **18,** 415–424 (2004).
2. Hollier, L. M., Leveno, K. J., Kelly, M. A., McIntire, D. D. & Cunningham, F. G. Maternal age and malformations in singleton births. *Obstet. Gynecol.* **96,** 701–706 (2000).
3. Kidd, S. A., Lancaster, P. A. & McCredie, R. M. The incidence of congenital heart defects in the first year of life. *J. Paediatr. Child Health* **29,** 344–349 (1993).
4. Materna-Kiryluk, A. *et al.* Parental age as a risk factor for isolated congenital malformations in a Polish population. *Paediatr. Perinat. Epidemiol.* **23,** 29–40 (2009).
5. Miller, A., Riehle-Colarusso, T., Siffel, C., Frias, J. L. & Correa, A. Maternal age and prevalence of isolated congenital heart defects in an urban area of the United States. *Am. J. Med. Genet. A* **155,** 2137–2145 (2011).
6. Pradat, P., Francannet, C., Harris, J. A. & Robert, E. The epidemiology of cardiovascular defects, part I: a study based on data from three large registries of congenital malformations. *Pediatr. Cardiol.* **24,** 195–221 (2003).
7. Reefhuis, J. & Honein, M. A. Maternal age and non-chromosomal birth defects, Atlanta—1968–2000: teenager or thirty-something, who is at risk? *Birth Defects Res. A Clin. Mol. Teratol.* **70,** 572–579 (2004).
8. Winston, J. B. *et al.* Complex trait analysis of ventricular septal defects caused by *Nkx2-5* mutation. *Circ Cardiovasc Genet* **5,** 293–300 (2012).
9. MRC Vitamin Study Research Group. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. *Lancet* **338,** 131–137 (1991).
10. Winston, J. B. *et al.* Heterogeneity of genetic modifiers ensures normal cardiac development. *Circulation* **121,** 1313–1321 (2010).
11. Schott, J. J. *et al.* Congenital heart disease caused by mutations in the transcription factor *NKX2-5*. *Science* **281,** 108–111 (1998).
12. Benson, D. W. *et al.* Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *J. Clin. Invest.* **104,** 1567–1573 (1999).
13. Nadeau, J. H. Modifier genes in mice and humans. *Nature Rev. Genet.* **2,** 165–174 (2001).
14. Jenkins, K. J. *et al.* Noninherited risk factors and congenital cardiovascular defects: current knowledge: a scientific statement from the American Heart Association Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation* **115,** 2995–3014 (2007).
15. Baird, P. A., Sadovnick, A. D. & Yee, I. M. Maternal age and birth defects: a population study. *Lancet* **337,** 527–530 (1991).
16. Loane, M., Dolk, H. & Morris, J. K. Maternal age-specific risk of non-chromosomal anomalies. *BJOG* **116,** 1111–1119 (2009).
17. Burrage, L. C. *et al.* Genetic resistance to diet-induced obesity in chromosome substitution strains of mice. *Mamm. Genome* **21,** 115–129 (2010).
18. Singer, J. B. *et al.* Genetic dissection of complex traits with chromosome substitution strains of mice. *Science* **304,** 445–448 (2004).
19. Allen, D. L. *et al.* Cardiac and skeletal muscle adaptations to voluntary wheel running in the mouse. *J. Appl. Physiol.* **90,** 1900–1908 (2001).
20. Li, M. *et al.* Detecting maternal–fetal genotype interactions associated with conotruncal heart defects: a haplotype-based analysis with penalized logistic regression. *Genet. Epidemiol.* **38,** 198–208 (2014).
21. Bye, A. *et al.* Serum levels of choline-containing compounds are associated with aerobic fitness level: the HUNT-study. *PLoS ONE* **7,** e42330 (2012).
22. Chorell, E., Svensson, M. B., Moritz, T. & Antti, H. Physical fitness level is reflected by alterations in the human plasma metabolome. *Mol. Biosyst.* **8,** 1187–1196 (2012).
23. Krug, S. *et al.* The dynamic range of the human metabolome revealed by challenges. *FASEB J.* **26,** 2607–2619 (2012).
24. Lewis, G. D. *et al.* Metabolic signatures of exercise in human plasma. *Sci. Transl. Med.* **2,** 33ra37 (2010).
25. Lustgarten, M. S. *et al.* Identification of serum analytes and metabolites associated with aerobic capacity. *Eur. J. Appl. Physiol.* **113,** 1311–1320 (2013).
26. Mukherjee, K. *et al.* Whole blood transcriptomics and urinary metabolomics to define adaptive biochemical pathways of high-intensity exercise in 50–60 year old masters athletes. *PLoS ONE* **9,** e92031 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements C.E.S. was supported by a Ruth L. Kirschstein National Research Service Award from the Developmental Cardiology and Pulmonary Training Program (National Institutes of Health (NIH) T32 HL007873). P.Y.J. is an Established Investigator of the American Heart Association and the Lawrence J. & Florence A. DeGeorge Charitable Trust. This work was supported by the Children's Discovery Institute of Washington University and St Louis Children's Hospital, the Children's Heart Foundation, and the NIH (R01 HL105857). The Washington University Digestive Diseases Research Core Center provided histology services and is supported by the NIH (P30 DK52574). MRI studies were performed in the Washington University Diabetes Research Center, which is supported by the NIH (P30 DK020579). We thank J. Magee, J. Rubin, D. Rudnick and A. Schwartz for comments.

Author Contributions C.E.S., D.B.W. and P.Y.J. designed experiments. C.E.S., S.D.R., R.A.M., M.T.D., H.L., A.K.H., A.A.P., M.M.G. and P.Y.J. executed experiments. C.E.S. and P.Y.J. interpreted experiments. C.E.S. and P.Y.J. wrote the manuscript. D.B.W. critically reviewed the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.Y.J. (jay_p@kids.wustl.edu).

METHODS

Mouse strains. The C57BL/6N and FVB/N inbred strains were purchased from Charles River, and the A/J strain from the Jackson Laboratory. *Nkx2-5*$^{+/-}$ mice were maintained in the C57BL/6N background; *Nkx2-5*$^{+/-}$ males were crossed to FVB/N or A/J females to produce first generation (F1) hybrids[10,27]. *Nkx2-5*$^{+/-}$ F1 mice were then crossed to produce the F2 offspring (Extended Data Fig. 1). No randomization was necessary among genetically identical mice. The animal studies committee at Washington University School of Medicine approved the experiments.

Preparation and phenotyping of newborn mouse hearts. Newborn mouse pups were collected within hours of birth. Their hearts were fixed in formalin, and the pups were genotyped for the *Nkx2-5* knockout allele. *Nkx2-5*$^{+/-}$ hearts were embedded in paraffin, serially sectioned at 6 μm thickness and stained with haematoxylin and eosin. The sections were examined for heart defects as previously described[10]. Phenotyping was performed by individuals blinded to the age of a pup's mother or, in the transplantation experiment, blinded to the age of the ovary as well. Every heart that was phenotyped is included in statistical analyses.

Ovarian transplantation. Mice were anaesthetized with a cocktail (0.5–0.7 ml kg^{-1}) of 3:3:1 ketamine (100 mg ml^{-1}), xylazine (20 mg ml^{-1}) and acepromazine (10 mg ml^{-1}) via intraperitoneal injection. A small incision was made in the dorsolateral flank to access the ovarian bursa. The ovary was removed from the donor and then transplanted into the recipient. The native, contralateral ovary of the recipient was left in place, while the oviduct was ligated. Transplants were performed between C57BL/6N × FVB/N F1 females that were either old (range 244–377 days old, mean 317 days old) or young (range 30–81 days old, mean 48 days old). The recipient females were allowed to recover for 3 weeks before breeding commenced. Pregnant females were noted to carry their fetuses in the uterine horn on the same side as the transplanted ovary.

High-fat diet. Weanling *Nkx2-5*$^{+/-}$ C57BL/6N × FVB/N F1 females were placed on a high-fat diet for 4 weeks before the onset of breeding. The mice remained on the diet while breeding. Calories in the diet derived from 59% fat, 25% carbohydrate and 15% protein (AIN-76A w/58% Fat Energy/Sucrose/Red, catalogue no. 1810835, Test Diet). Calories in the normal, control diet derived from 13% fat, 62% carbohydrate and 25% protein (Pico Rodent Diet 20, catalogue no. 0007688, Lab Diet).

Exercise. Running wheels were placed in breeding cages when the females were either 4 weeks old (early onset group) or 8 months old (late onset group). The mice ran *ad libitum* for the remainder of their reproductive lives.

Intraperitoneal glucose tolerance testing. Mice were fasted overnight (14 h) on paper bedding before glucose challenge (2 g kg^{-1} intraperitoneal). Blood was obtained from the tail vein for glucometry. Samples were measured with a Bayer Contour TS glucometer before and at 10, 20, 30, 60, 90 and 120 min after injection.

Lean and fat body mass quantification. Lean and fat body mass was measured in live mice by quantitative MRI on an EchoMRI 3-in-1 instrument (Echo Medical Systems). Fat body mass measurements were calibrated against canola oil standards.

Statistics. The maternal-age-associated risk of a defect within an experimental condition was calculated by logistic regression analysis. The phenotype of a pup, for example, VSD present or normal, was the dependent variable, and the mother's age in months on the pup's birthdate was the independent variable. The data were fit to the inverse of the logistic function or logit using the Generalized Linear Model in R (http://www.r-project.org/). Fitting yielded an estimate of the maternal age coefficient, from which the odds ratio is calculated. Odds ratios are reported with the 95% confidence interval.

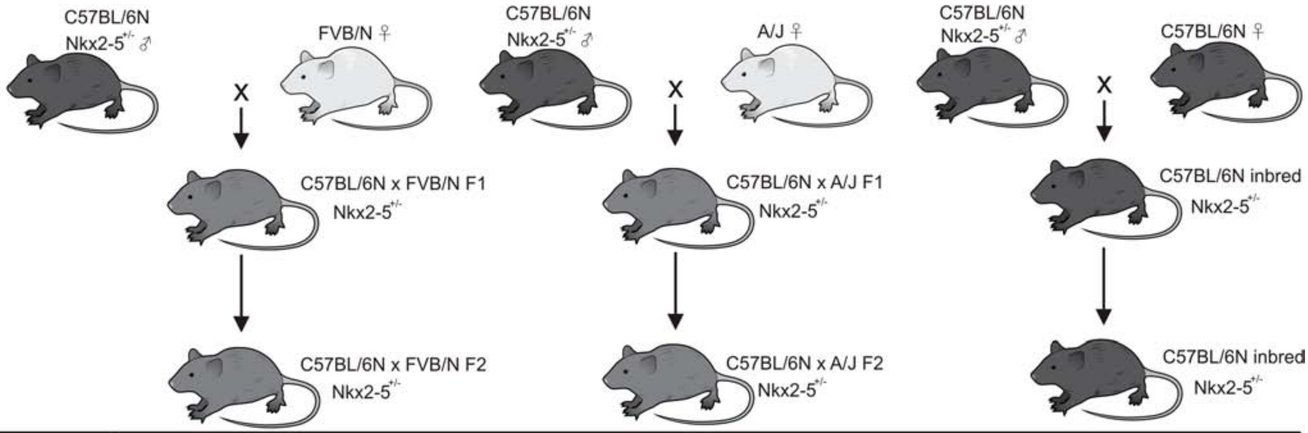
For the reciprocal ovarian transplant and late onset exercise experiments, the expected incidence of a defect was calculated from the estimates of the coefficients in the logit for the C57BL/6N × FVB/N sedentary control group and either the age of the mother or ovary for every pup. The probability of a defect, as given by the logistic function, was calculated for each observed pup. The sum of probabilities across all the pups was the expected number for the experimental group, given the hypothesized or null model. The expected and observed numbers of defects were compared in a chi-squared goodness-of-fit test. To assess the effect of other variables, such as genetic background, diet or the interaction of either with age, the variables were included in the logistic function as covariates in addition to maternal age. The significance of a covariate was determined using the Generalized Linear Model in R.

Two-sided *t*-tests were performed for blood glucose and body composition measurements.

Experimental sample sizes in the ovarian transplantation, high-fat diet, and exercise experiments and analysis of the C57BL/6N strain were planned based on the incidences of VSD in the offspring of young and old C57BL/6N × FVB/N F1 mothers. Every heart collected for an experimental condition was included in the statistical analysis unless poor histology precluded a diagnosis.

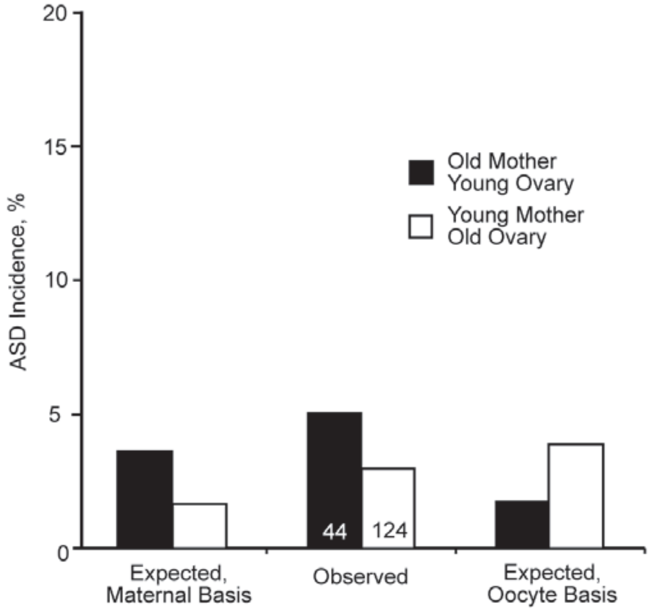
Results are reported as mean ± s.e.m. except for maternal age, which is reported as mean ± s.d. Statistical significance was defined as $P < 0.05$.

27. Tanaka, M., Chen, Z., Bartunkova, S., Yamasaki, N. & Izumo, S. The cardiac homeobox gene *Csx/Nkx2.5* lies genetically upstream of multiple genes essential for heart development. *Development* **126,** 1269–1280 (1999).

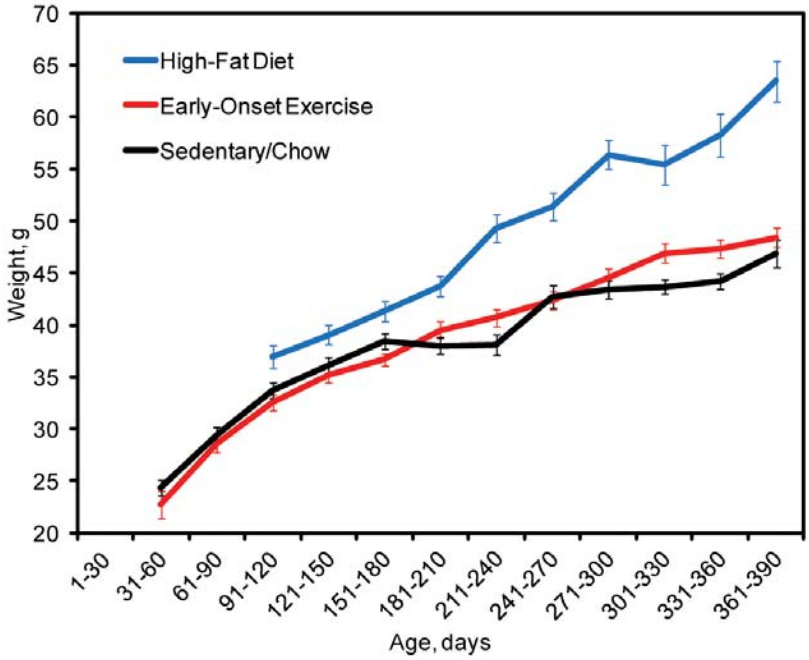


Experimental Condition	Mothers, N		
	C57BL/6N x FVB/N F1	C57BL/6N x A/J F1	C57BL/6N inbred
Sedentary/Chow	156	66	31
High-Fat Diet	19		
Early-Onset Exercise	15		
Late-Onset Exercise	12		

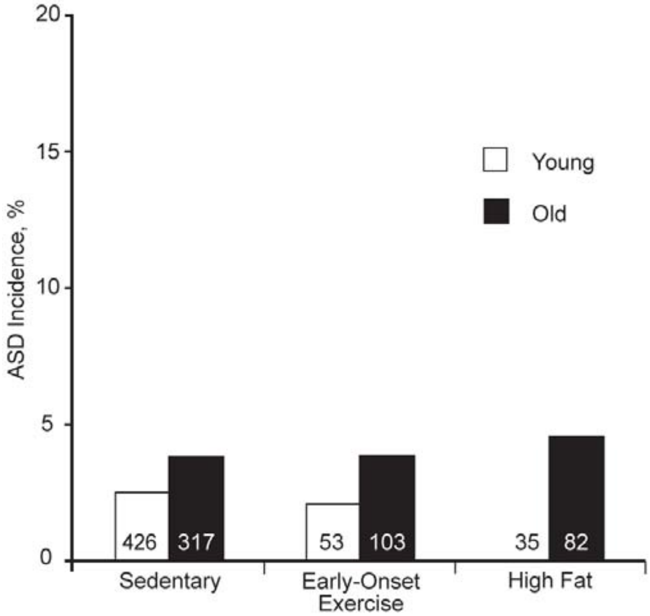
Extended Data Figure 1 | Breeding scheme and experimental conditions. *Nkx2-5*$^{+/-}$ offspring from several maternal genetic backgrounds and experimental conditions were phenotyped. *Nkx2-5*$^{+/-}$ C57BL/6N males were crossed to FVB/N or A/J females to produce F1 hybrids. The cross to a C57BL/6N female maintains the inbred strain. *Nkx2-5*$^{+/-}$ F1 hybrids were intercrossed to produce the F2 progeny. The hearts of newborn *Nkx2-5*$^{+/-}$ F2 pups were phenotyped to calculate the incidence of a defect and the effect of maternal age. C57BL/6N × FVB/N F1 hybrid mothers were bred in either sedentary/chow, high-fat diet, early or late onset exercise conditions. C57BL/6N × A/J F1 hybrid and C57BL/6N inbred mothers were studied only in the sedentary/chow condition. The number of mothers in each cross and experimental condition that were used to produce pups in this study are shown.



Extended Data Figure 2 | Incidences of ASD in the reciprocal ovarian transplant experiment. The relative incidences of ASD in the reciprocal ovarian transplant experiment are consistent with a maternal basis of the age-associated effect. The differences that were significant in the VSD data, however, are not significant here because of the lower incidence of ASD, as depicted by the *y*-axis drawn on a scale comparable to that for VSD. The total number of pups in each group is shown.



Extended Data Figure 3 | Growth charts for C57BL/6N × FVB/N F1 mothers under sedentary, early onset exercise and high-fat diet conditions. C57BL/6N × FVB/N F1 mothers on a high-fat diet develop marked obesity as they age. They weigh substantially more than mothers in the sedentary or early onset exercise groups. Mothers in the latter two groups weigh the same.



Extended Data Figure 4 | Incidences of ASD in the offspring of C57BL/6N × FVB/N mothers. Maternal age may affect the risk of ASD, but the lower incidence of ASD and other defects that are less common than VSD preclude firm statistical conclusions. For example, the incidences of ASD are shown for the *Nkx2-5*$^{+/-}$ offspring of young and old C57BL/6N × FVB/N mothers in the sedentary, early onset exercise, and high-fat diet conditions. The *y*-axis is drawn on a scale comparable to that for VSD incidence. ASD incidences are higher, but not significantly, among the offspring of old mothers compared to young mothers. The incidences are not significantly different between experimental conditions. The total number of pups in each group is shown.

doi:10.1038/nature14359

Repeated ER–endosome contacts promote endosome translocation and neurite outgrowth

Camilla Raiborg[1,2], Eva M. Wenzel[1,2], Nina M. Pedersen[1,2], Hallvard Olsvik[3], Kay O. Schink[1,2], Sebastian W. Schultz[1,2], Marina Vietri[1,2], Veronica Nisi[4], Cecilia Bucci[4], Andreas Brech[1,2], Terje Johansen[3] & Harald Stenmark[1,2]

The main organelles of the secretory and endocytic pathways—the endoplasmic reticulum (ER) and endosomes, respectively—are connected through contact sites whose numbers increase as endosomes mature[1–3]. One function of such sites is to enable dephosphorylation of the cytosolic tails of endosomal signalling receptors by an ER-associated phosphatase[4], whereas others serve to negatively control the association of endosomes with the minus-end-directed microtubule motor dynein[5] or mediate endosome fission[6]. Cholesterol transfer and Ca^{2+} exchange have been proposed as additional functions of such sites[2,3]. However, the compositions, activities and regulations of ER–endosome contact sites remain incompletely understood. Here we show in human and rat cell lines that protrudin, an ER protein that promotes protrusion and neurite outgrowth[7], forms contact sites with late endosomes (LEs) via coincident detection of the small GTPase RAB7 and phosphatidylinositol 3-phosphate (PtdIns(3)P). These contact sites mediate transfer of the microtubule motor kinesin 1 from protrudin to the motor adaptor FYCO1 on LEs. Repeated LE–ER contacts promote microtubule-dependent translocation of LEs to the cell periphery and subsequent synaptotagmin-VII-dependent fusion with the plasma membrane. Such fusion induces outgrowth of protrusions and neurites, which requires the abilities of protrudin and FYCO1 to interact with LEs and kinesin 1. Thus, protrudin-containing ER–LE contact sites are platforms for kinesin-1 loading onto LEs, and kinesin-1-mediated translocation of LEs to the plasma membrane, fuelled by repeated ER contacts, promotes protrusion and neurite outgrowth.

The unusual presence of a PtdIns(3)P-binding FYVE domain in the ER protein protrudin[7–10] begged the question as to whether this protein interacts with endosomes, the major PtdIns(3)P-containing organelles in cells[11]. Confocal microscopy of HeLa cells transfected with green fluorescent protein (GFP)-tagged protrudin showed that LAMP1-positive LEs and lysosomes colocalized extensively with GFP–protrudin-positive ER (Fig. 1a, top). Co-transfection with a GTPase-defective mutant of the LE-associated GTPase RAB7 (ref. 12), RAB7(Q67L) (ref. 13), strongly increased the recruitment of GFP–protrudin to LEs (Fig. 1b and Extended Data Fig. 1a), whereas a GTPase-defective mutant of the early endosomal

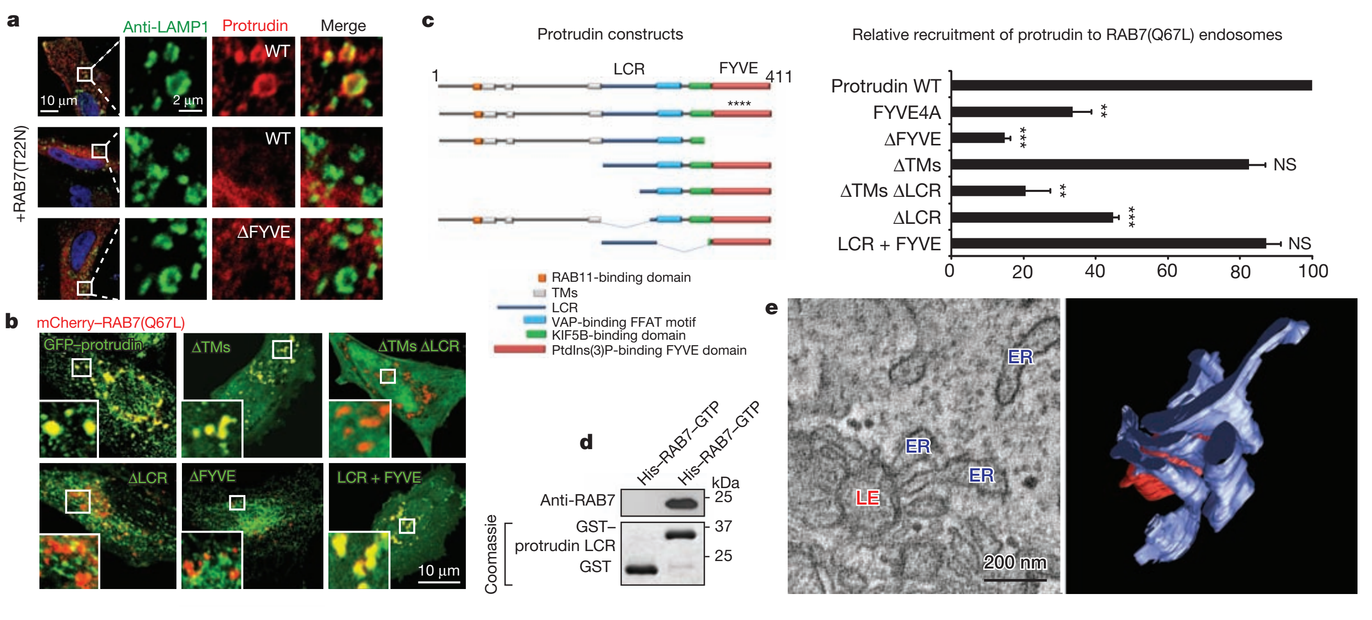


Figure 1 | Protrudin forms RAB7- and PtdIns(3)P-dependent ER–LE contact sites. **a**, Protrudin recruitment to LAMP1-positive endosomes depends on functional RAB7 and an intact FYVE domain. Confocal images are representative of at least 10 captures. WT, wild type. **b**, The FYVE and LCR domains of protrudin mediate its recruitment to RAB7(Q67L)-positive endosomes (representative of at least 10 confocal captures and 240 wide-field images quantified in **c**). TMs, transmembrane domains. **c**, Quantification of the relative recruitment of protrudin to RAB7(Q67L). Error bars denote ± standard error of the mean (s.e.m.). Protrudin LCR + FYVE: 3 experiments, 1,463 cells; protrudin ΔLCR: 3 experiments, 4,792 cells; protrudin ΔTMs ΔLCR: 3 experiments, 3,499 cells; protrudin ΔTMs: 3 experiments, 1,716 cells; protrudin ΔFYVE: 6 experiments, 6,121 cells; protrudin FYVE4A: 6 experiments, 5,583 cells. $^{**}P < 0.01$, $^{***}P < 0.001$ (one-sample t-test). NS, non significant. **d**, *In vitro* pull-down of a GST–protrudin–LCR construct (residues 208–274) shows direct binding to His–RAB7 loaded with GTP. **e**, Tomography from a HeLa cell transiently expressing Myc–protrudin and GFP–RAB7(Q67L) showing how ER wraps around a late endosome. Representative of two tomograms.

[1]Centre for Cancer Biomedicine, Faculty of Medicine, University of Oslo, Montebello, N-0379 Oslo, Norway. [2]Department of Molecular Cell Biology, Institute for Cancer Research, Oslo University Hospital, Montebello, N-0379 Oslo, Norway. [3]Institute of Medical Biology, University of Tromsø — The Arctic University of Norway, N-9037 Tromsø, Norway. [4]Department of Biological and Environmental Sciences and Technologies (DiSTeBA), University of Salento, Via Provinciale Monteroni 165, 73100 Lecce, Italy.

GTPase RAB5 (ref. 14) did not recruit protrudin (Extended Data Fig. 1b). Deletion and point mutagenesis showed that both the FYVE domain and the low-complexity region (LCR) of protrudin were necessary for protrudin recruitment to RAB7(Q67L) endosomes, and a construct consisting of only these two domains was efficiently recruited (Fig. 1b, c). A GDP-bound mutant of RAB7, RAB7(T22N), failed to recruit protrudin (Fig. 1a, middle, and Extended Data Fig. 1c) and GFP-trap pull-downs from cell lysates indicated a specific interaction between Myc-tagged protrudin and GFP–RAB7(Q67L) (Extended Data Fig. 1d). This interaction was retained in a carboxy-terminal protrudin mutant containing the LCR but not in a mutant that lacked the LCR (Extended Data Fig. 1e). Indeed, glutathione *S*-transferase (GST) pull-down assays with purified proteins showed that GST-tagged LCR, but not GST alone, interacted directly with GTP- (but not GDP-) loaded RAB7, and that there was no interaction with RAB5–GTP (Fig. 1d and Extended Data Fig. 1f). Thus, protrudin is a novel RAB7 effector through its LCR.

Confocal microscopy of transfected HeLa cells showed that Myc–protrudin colocalized specifically with PtdIns(3)P-positive LEs containing RAB7 (Extended Data Fig. 1g), as visualized with a 2×FYVE probe expressed at a low level[11]. Consistent with this, deletion of the FYVE domain prevented recruitment of protrudin to LAMP1-positive LEs (Fig. 1a, bottom). High-level expression of the 2×FYVE probe is known to sequester PtdIns(3)P[15], and GFP–2×FYVE was found to prevent accumulation of Myc–protrudin on RAB7(Q67L) endosomes, as opposed to the 2×FYVE(C215S) mutant, which does not bind PtdIns(3)P (Extended Data Fig. 1h). Taken together, our results indicate that protrudin is recruited to LEs through coincident detection of RAB7–GTP via the LCR and PtdIns(3)P via the FYVE domain.

Recruitment of protrudin to RAB7(Q67L) LEs was accompanied by ER recruitment, as revealed by confocal microscopy using an antibody against the reticular ER marker VAP-A[16] (Extended Data Fig. 2a). Transmission electron microscopy showed that the ER formed multiple contacts with LEs in cells co-transfected with protrudin and RAB7(Q67L) (Extended Data Fig. 2b, c). Electron tomography revealed that the contacts were formed both at the tips and the body of ER tubules, reorganizing the overall ER structure (Fig. 1e and Supplementary Video 1). Correlative light and electron microscopy (CLEM) showed a complete sequestration of ER membranes around LEs in cells expressing high levels of protrudin and RAB7(Q67L) (Extended Data Fig. 3). These data indicate that protrudin forms contact sites with LEs, and that excessive contact site formation via overexpression of protrudin and RAB7(Q67L) causes a massive redistribution of ER around LEs.

Surprisingly, overexpression of protrudin was found to cause a strong accumulation of LAMP1-positive LEs in the cell periphery, and such translocation required the protrudin LCR and FYVE domains and functional RAB7 (Fig. 2a). In contrast, the localization of early and recycling endosomes was not detectably changed by protrudin overexpression (Extended Data Fig. 4a). Another PtdIns(3)P- and RAB7-binding LE protein, FYCO1, has previously been shown to promote localization of LEs to the cell periphery[17], and FYCO1 was indeed found to colocalize with protrudin on LAMP1-positive LEs (Fig. 2b). Endogenous FYCO1 accumulated strongly on RAB7(Q67L) LEs, where it was surrounded by protrudin (Extended Data Fig. 4b), and a GFP-trap assay revealed the presence of GFP–FYCO1 and Myc–protrudin in the same complex (Extended Data Fig. 4c). Live tracking of individual FYCO1 LEs showed a strong net movement towards the cell periphery in protrudin-overexpressing

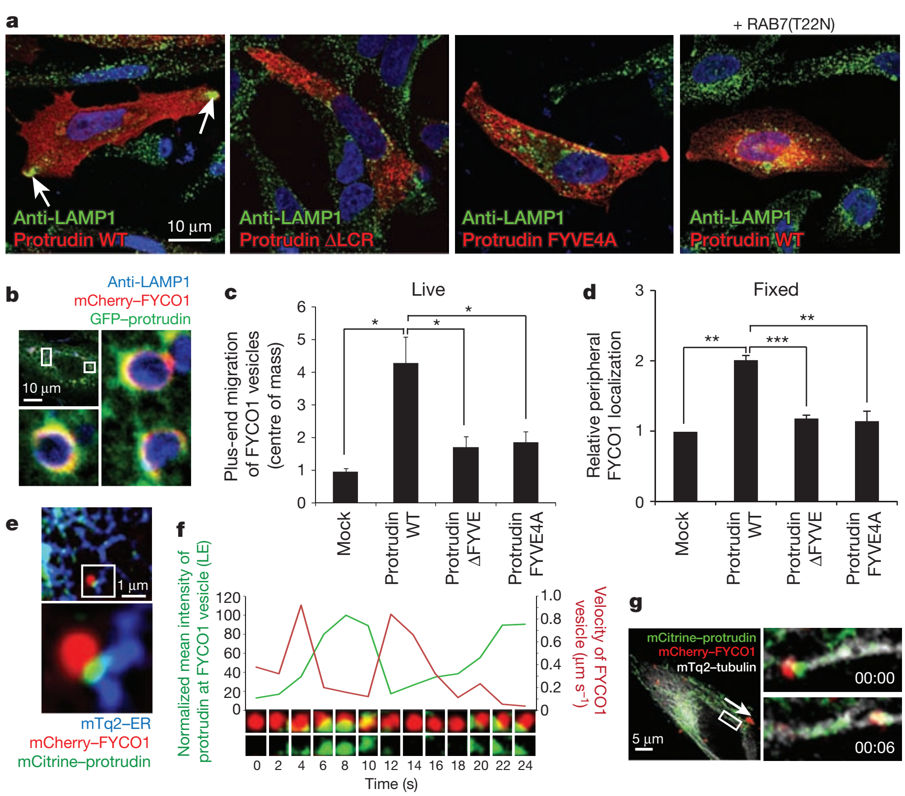


Figure 2 | Protrudin forms repeated contact sites with LEs and mediates their translocation to the cell periphery. **a**, LEs accumulate close to the plasma membrane in protrudin wild type (WT; arrows), but not in protrudin(FYVE4A)-, protrudin(ΔLCR)- or RAB7(T22N)-overexpressing cells. Confocal images are representative of at least 10 captures. **b**, GFP–protrudin-positive ER wraps around mCherry–FYCO1-decorated LAMP1-positive late endosomes. Confocal images are representative of at least 10 captures. **c**, Quantification of directional trafficking of mCherry–FYCO1 vesicles (Extended Data Fig. 4d). Centre of mass >0 means plus-end migration. Error bars denote ± s.e.m. GFP: 3 experiments (12 cells, 62 vesicles); GFP–protrudin: 3 experiments (10 cells, 58 vesicles); GFP–protrudin(ΔFYVE): 4 experiments (15 cells, 89 vesicles); GFP–protrudin(FYVE4A): 4 experiments (16 cells, 81 vesicles). $*P < 0.05$ (unpaired *t*-test). **d**, Quantification of the ratio of peripheral versus perinuclear localization of mCherry–FYCO1 upon overexpression of the indicated constructs. Error bars denote ± s.e.m. from 3 experiments. GFP: 1,296 cells; GFP–protrudin: 1,208 cells; GFP–protrudin(ΔFYVE): 1,048 cells; GFP–protrudin(FYVE4A): 918 cells. $**P < 0.01$, $***P < 0.001$. Mock/protrudin, one-sample *t*-test. Protrudin/ΔFYVE/FYVE4A, unpaired *t*-test. **e**, Still image from Supplementary Video 3 showing an mCherry–FYCO1 LE making contact with mCitrine–protrudin-positive ER. Representative of 4 videos. **f**, An mCherry–FYCO1 vesicle repeatedly contacts GFP–protrudin when its speed slows down. Images are magnifications from the vesicle in Extended Data Fig. 6a, representative of 31 videos. **g**, Still images from Supplementary Video 5 showing how an mCherry–FYCO1 vesicle moves along a microtubule (seconds). Representative of 12 videos.

cells (Fig. 2c, Extended Data Fig. 4d and Supplementary Video 2). This effect was dependent on the FYVE domain of protrudin as shown in both live (Fig. 2c) and fixed cells (Fig. 2d).

Conversely, when HeLa or retinal pigment epithelial cells (RPE1) were depleted of protrudin using short interfering RNA (siRNA), LAMP1- or FYCO1-positive LEs showed a stronger perinuclear distribution than in control cells (Extended Data Fig. 5a–d). Similar results were obtained in cells depleted of kinesin light chain 1 (KLC1) (Extended Data Fig. 5a, c). The knockdown phenotype could be rescued by low-level expression of siRNA-resistant protrudin (Extended Data Fig. 5e). LAMP1-positive LEs also clustered perinuclearly in cells depleted of FYCO1 by siRNA (Extended Data Fig. 5f). We conclude that protrudin mediates translocation of LEs to the cell periphery in a FYCO1-dependent manner.

Contacts between mCherry–FYCO1-positive LEs and protrudin-containing ER could be visualized by live imaging (Fig. 2e and Supplementary Video 3). FYCO1 LEs were found to associate repeatedly with protrudin during their translocation towards the plasma membrane, and showed slow random motility when associated with protrudin and fast directional motility after leaving protrudin (Fig. 2f, Extended Data Fig. 6a–d and Supplementary Video 4). Accordingly, we found a pronounced inverse correlation between LE speed and protrudin association (-0.6 ± 0.1 standard deviation (s.d.)) measured on 31 LEs using Spearman's correlation coefficient). Fast directional movement of mCherry–FYCO1 LEs occurred along microtubules as visualized with mTq2-tagged tubulin (Fig. 2g and Supplementary Video 5) with an average maximum velocity of $0.9\ \mu m\ s^{-1} \pm 0.3$ s.d., in line with previous reports of microtubule-dependent transport[18]. These results indicate that FYCO1 LEs associate repeatedly with different protrudin-containing ER domains, and that fast directional movements between such domains occur along microtubules in the plus-end direction.

ER is intertwined with microtubules[1], and protrudin has previously been shown to bind the KIF5 heavy chain of the plus-end-directed microtubule motor kinesin 1 (ref. 19). Interestingly, bead capture assays indicated that the middle part of FYCO1 (residues 585–1233) interacts directly with the KLC2 light chain of kinesin 1 (Fig. 3a and Extended Data Fig. 7a, b), and GFP–FYCO1 was found in complex with KIF5B in a GFP-trap assay (Fig. 3b), suggesting that FYCO1 functions as a kinesin-1 adaptor on LEs. Accordingly, FYCO1 vesicles were positive for the endogenous KIF5B and KLC1 chains of kinesin 1 (Extended Data Fig. 7c, d). Residues 735–773 of FYCO1 were found to be necessary for its kinesin-1 binding (Fig. 3c and Extended Data Fig. 7b, c), and FYCO1(Δ735–773)-positive LEs failed to translocate to the cell periphery (Extended Data Fig. 7e). Moreover, siRNA-mediated depletion of RAB7, which led to dissociation of FYCO1 from LEs (data not shown), strongly prevented pull-down of KIF5B with GFP–FYCO1 (Fig. 3d and Extended Data Fig. 7f). Thus, FYCO1 interacts with kinesin 1 in a RAB7-dependent manner.

Interestingly, Myc-tagged protrudin, but not a point mutant defective in phosphoinositide binding, protrudin(FYVE4A)[20], caused an increase in kinesin 1 associated with FYCO1 LEs (Extended Data Fig. 7d, g). GFP-trap experiments showed that Myc–protrudin promoted association of KIF5B with GFP–FYCO1 (Fig. 3e and Extended Data Fig. 7h). In contrast, a mutant protrudin construct lacking the KIF5B-binding site[19] failed to promote association of KIF5B with GFP–FYCO1 and even

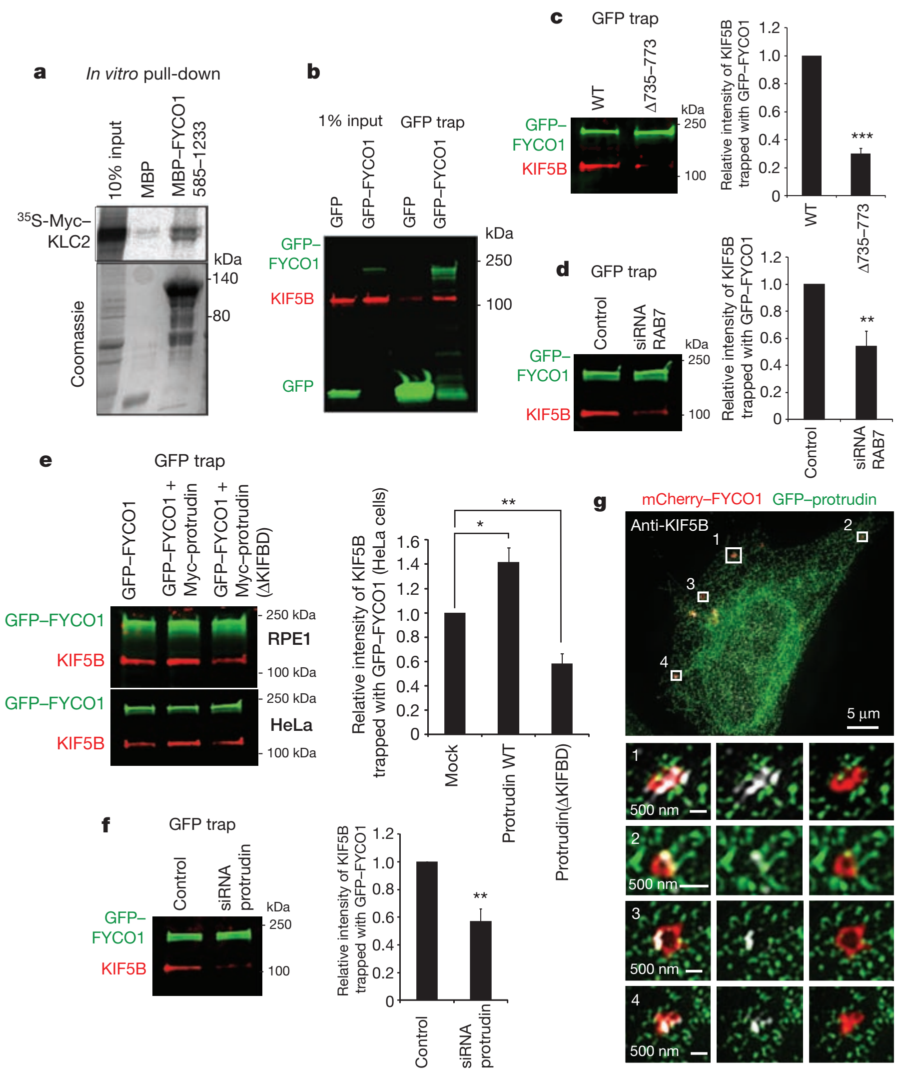


Figure 3 | Protrudin loads kinesin 1 onto FYCO1-positive LEs. **a**, *In vitro* translated ^{35}S-Myc–KLC2 interacts with MBP–FYCO1$^{585–1233}$ beads. **b**, KIF5B is pulled down with GFP–FYCO1 in a GFP-trap assay. **c**, Residues 735–773 of FYCO1 are responsible for its KIF5B association. Quantification of the relative intensity of KIF5B trapped with GFP–FYCO1. WT, wild type. Error bar denotes ± s.e.m. from 5 experiments. ***$P < 0.001$ (one-sample *t*-test). **d**, Less KIF5B is pulled down with GFP–FYCO1 in cells depleted of RAB7 by siRNA. Error bar denotes ± s.e.m. from 9 experiments. **$P < 0.01$ (one-sample *t*-test). **e**, Protrudin regulates the level of KIF5B pulled down with GFP–FYCO1. Error bars denote ± s.e.m. Protrudin wild type: 4 experiments; protrudin(ΔKIFBD): 6 experiments. *$P < 0.05$, **$P < 0.01$ (one-sample *t*-test). **f**, Less KIF5B is pulled down with GFP–FYCO1 in cells depleted of protrudin. Error bar denotes ± s.e.m. from 6 experiments. **$P < 0.01$ (one-sample *t*-test). **g**, Structured illumination microscopy (SIM) image showing how endogenous KIF5B localizes to mCherry–FYCO1-positive vesicles in a GFP–protrudin-overexpressing cell. Representative of 14 images.

reduced KIF5B–FYCO1 association as compared to mock-transfected cells (Fig. 3e and Extended Data Fig. 7h). Moreover, siRNA-mediated depletion of protrudin decreased the amount of KIF5B pulled down with GFP–FYCO1 (Fig. 3f and Extended Data Fig. 7i), in line with the perinuclear localization of GFP–FYCO1 endosomes in protrudin-depleted cells (Extended Data Fig. 5d). Thus, the kinesin-1-binding ability of protrudin mediates kinesin-1 loading onto FYCO1, and such loading is required for LE translocation. We conclude that ER-associated protrudin promotes LE translocation by loading of kinesin 1 onto FYCO1 on LEs, presumably by a direct handover mechanism, and that this can only happen when ER–LE contacts are firmly established and FYCO1 is attached to LEs. This was further supported by structured illumination microscopy showing that KIF5B closely colocalizes with FYCO1 LEs in protrudin-overexpressing cells (Fig. 3g).

To investigate a potential role of the protrudin–FYCO1 axis for promoting protrusions[7], we used RPE1 cells in which GFP–protrudin was found to induce long protrusions in a proportion of cells (Extended Data Fig. 8a). Interestingly, when mCherry–FYCO1 was co-expressed with GFP–protrudin in these cells, FYCO1 endosomes were typically observed in the tips of protrusions (Extended Data Fig. 8a), and the co-expression strongly increased the percentage of cells containing protrusions as observed with both fixed and live cells. This required protrudin to have a functional FYVE domain (Extended Data Fig. 8b, c and Supplementary Video 6). Conversely, when FYCO1 or RAB7 was depleted using siRNA, protrudin-mediated protrusion formation strongly decreased (Extended Data Fig. 8d, e). Thus, protrudin and FYCO1 cooperatively promote protrusions.

The neuroendocrine cell line PC12 is a well-established model for studying nerve growth factor (NGF)-induced neurite outgrowth[7,21], and GFP–protrudin induced both NGF-dependent and -independent neurites in such cells (Extended Data Fig. 9a), in line with previous reports[7,8]. Confocal microscopy showed that FYCO1 endosomes were frequently located in the neurite tips (Fig. 4a and Supplementary Video 7). Interestingly, neurite outgrowth not only required the functional FYVE domain of protrudin[20] but also the RAB7-binding LCR (Fig. 4b). Moreover, co-transfection with mCherry–FYCO1 further stimulated neurite outgrowth whereas this was not the case with the FYCO1 mutant (Δ735–773) lacking the kinesin-1-binding domain (Fig. 4b and Extended Data Fig. 9b). Importantly, while depletion of protrudin inhibited NGF-stimulated neurite outgrowth as shown previously[7,22], this was also the case when PC12 cells were depleted of FYCO1 or RAB7 (Fig. 4c, d and Extended Data Fig. 9c). Collectively, our data indicate that protrudin and FYCO1 cooperate to form protrusions and neurites in a RAB7-, PtdIns(3)P- and kinesin-1-dependent manner.

Previous studies have identified a role for synaptotagmin VII (SYT7)-mediated fusion of lysosomes with the plasma membrane in neurite outgrowth[23]. Indeed, depletion of SYT7 strongly inhibited protrudin- and FYCO1-mediated protrusion formation in RPE1 cells (Extended Data Fig. 9d), suggesting a role for SYT7-dependent fusion of LEs with the plasma membrane in this process. A LAMP1 construct with a GFP

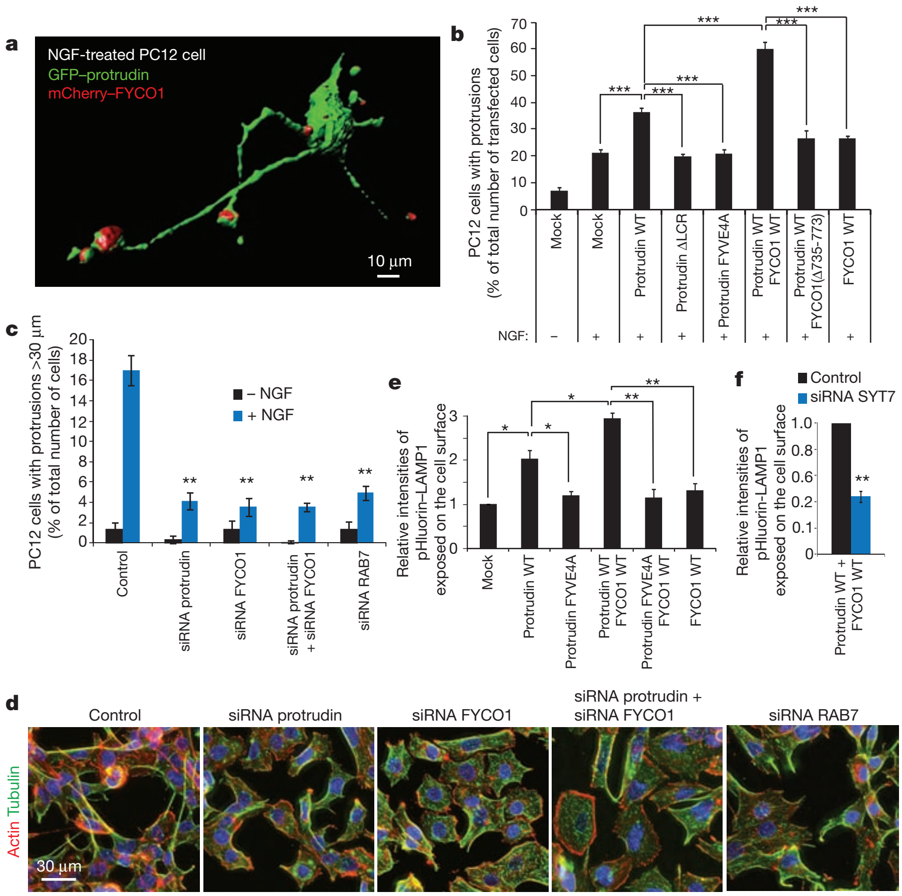


Figure 4 | Protrudin and FYCO1 cooperatively induce neurite outgrowth by facilitating LE fusion with the plasma membrane. **a**, Imaris surface view of a confocal z-stack showing GFP–protrudin and mCherry–FYCO1 in a PC12 cell. Representative of 3 confocal z-stacks and at least 80 wide-field images. **b**, Protrudin-mediated neurite outgrowth in PC12 cells is dependent on protrudin–endosome contact and FYCO1. Error bars denote ± s.e.m. The following experiments were repeated 4 times: GFP + mCherry (−NGF): 313 cells; GFP + mCherry (+NGF): 284 cells; GFP–protrudin + mCherry: 1,190 cells; GFP–protrudin(ΔLCR) + mCherry: 291 cells, GFP–protrudin(FYVE4A) + mCherry: 186 cells. The following experiments were repeated 3 times: GFP–protrudin + mCherry–FYCO1: 497 cells; GFP–protrudin + mCherry–FYCO1(Δ735–773): 489 cells; GFP + mCherry–FYCO1: 297 cells. ***$P < 0.001$ (unpaired *t*-test). **c**, Neurite outgrowth is dependent on protrudin, FYCO1 and RAB7. Error bars denote ± s.e.m. of 3 experiments. The following experiments were carried out without NGF: control: 319 cells; siRNA protrudin: 262 cells; siRNA FYCO1: 231 cells; siRNA protrudin + FYCO1: 218 cells; siRNA RAB7: 224 cells. The following experiments were carried out with NGF: control: 1,202 cells; siRNA protrudin: 621 cells; siRNA FYCO1: 674 cells; siRNA protrudin + FYCO1: 639 cells; siRNA RAB7: 607 cells. **$P < 0.01$ (unpaired *t*-test). **d**, Images of PC12 cells quantified in **c**. Representative of at least 30 wide-field images per condition. **e**, Protrudin and FYCO1 increase fusion of pHluorin–LAMP1 endosomes with the plasma membrane. Error bars denote ± s.e.m. from 3 experiments. Mock: 68 cells; protrudin: 72 cells; protrudin(FYVE4A): 73 cells; protrudin + FYCO1: 94 cells; protrudin(FYVE4A) + FYCO1: 70 cells; FYCO1: 78 cells. *$P < 0.05$, **$P < 0.01$. Mock/protrudin, one-sample *t*-test. All others, unpaired *t*-test. **f**, Protrudin- and FYCO1-mediated LE–PM fusion is dependent on SYT7. Error bars denote ± s.e.m. from 3 experiments. Control: 63 cells; siRNA SYT7: 62 cells. **$P < 0.01$ (one-sample *t*-test).

variant on the luminal (extracellular) side was employed to monitor fusion of LEs with the plasma membrane, as detected with antibodies against extracellular GFP in unpermeabilized HeLa cells (Extended Data Fig. 9e). Confocal microscopy showed that overexpression of protrudin, but not protrudin(FYVE4A), caused a strong redistribution of LAMP1 to the cell surface, particularly in protrusion buds (Fig. 4e and Extended Data Fig. 9f), and co-transfection with FYCO1 increased the plasma membrane LAMP1 signal even further (Fig. 4e). These results correlate strongly with the cooperative effects of protrudin and FYCO1 in protrusion and neurite outgrowth (Fig. 4b and Extended Data Fig. 8b, c), and siRNA-mediated depletion of SYT7 indeed prevented protrudin- and FYCO1-mediated fusion of LEs with the plasma membrane (Fig. 4f). Thus, protrudin and FYCO1 cooperate to promote SYT7-dependent fusion of LEs with the plasma membrane, and this promotes protrusion outgrowth.

Our findings reveal an unanticipated role for ER–LE contact sites in the control of protrusion and neurite outgrowth (Extended Data Fig. 10). When contact sites are formed between ER and LEs via recognition of LE-associated PtdIns(3)P and RAB7–GTP by ER-bound protrudin, this enables transfer of kinesin 1 from protrudin to FYCO1 on LEs. Kinesin 1 can then propel the LE along microtubules in the plus direction towards the growing protrusion tips, where LE accumulation ultimately leads to SYT7-mediated fusion with the plasma membrane. This could provide lipids and proteins to the protrusion tip membrane that are favourable for protrusion extension[23].

Recent studies have shown that phosphoinositides can promote ER contact site formation with the Golgi and the plasma membrane[24–26]. We now show that a phosphoinositide and a RAB GTPase can cooperate in contact site formation. Both groups of molecules are subject to acute regulation[27–30], consistent with the highly dynamic formation and resolution of protrudin-containing ER–LE contacts.

Our observation that FYCO1 LEs contact the ER repeatedly during their translocation towards the plasma membrane suggests that they might require reloading with kinesin 1 to maintain their motility. Loading of LEs with kinesin-1 molecules immobilized on the ER instead of soluble in cytosol is likely to provide spatiotemporal control of motor loading. Formation of another type of ER–LE contacts, those mediated by ORP1L, leads to dissociation of the minus-end-directed motor dynein from LEs[5]. Thus, it will be interesting to learn whether protrudin- and ORP1L-containing ER–LE contact sites could form side by side to enable a tight coordination between dynein dissociation and kinesin-1 association.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 May 2014; accepted 27 February 2015.

1. Friedman, J. R., Dibenedetto, J. R., West, M., Rowland, A. A. & Voeltz, G. K. Endoplasmic reticulum–endosome contact increases as endosomes traffic and mature. *Mol. Biol. Cell* **24,** 1030–1040 (2013).
2. van der Kant, R. & Neefjes, J. Small regulators, major consequences—Ca^{2+} and cholesterol at the endosome–ER interface. *J. Cell Sci.* **127,** 929–938 (2014).
3. Hönscher, C. & Ungermann, C. A close-up view of membrane contact sites between the endoplasmic reticulum and the endolysosomal system: from yeast to man. *Crit. Rev. Biochem. Mol. Biol.* **49,** 262–268 (2014).
4. Eden, E. R., White, I. J., Tsapara, A. & Futter, C. E. Membrane contacts between endosomes and ER provide sites for PTP1B–epidermal growth factor receptor interaction. *Nature Cell Biol.* **12,** 267–272 (2010).
5. Rocha, N. *et al.* Cholesterol sensor ORP1L contacts the ER protein VAP to control Rab7–RILP–$p150^{Glued}$ and late endosome positioning. *J. Cell Biol.* **185,** 1209–1225 (2009).
6. Rowland, A. A., Chitwood, P. J., Phillips, M. J. & Voeltz, G. ER contact sites define the position and timing of endosome fission. *Cell* **159,** 1027–1041 (2014).
7. Shirane, M. & Nakayama, K. I. Protrudin induces neurite formation by directional membrane trafficking. *Science* **314,** 818–821 (2006).
8. Chang, J., Lee, S. & Blackstone, C. Protrudin binds atlastins and endoplasmic reticulum-shaping proteins and regulates network formation. *Proc. Natl Acad. Sci. USA* **110,** 14954–14959 (2013).
9. Pantakani, D. V., Czyzewska, M. M., Sikorska, A., Bodda, C. & Mannan, A. U. Oligomerization of ZFYVE27 (Protrudin) is necessary to promote neurite extension. *PLoS ONE* **6,** e29584 (2011).
10. Catimel, B. *et al.* The PI(3)P interactome from a colon cancer cell. *J. Proteomics* **82,** 35–51 (2013).
11. Gillooly, D. J. *et al.* Localization of phosphatidylinositol 3-phosphate in yeast and mammalian cells. *EMBO J.* **19,** 4577–4588 (2000).
12. Chavrier, P., Parton, R. G., Hauri, H. P., Simons, K. & Zerial, M. Localization of low molecular weight GTP binding proteins to exocytic and endocytic compartments. *Cell* **62,** 317–329 (1990).
13. Cantalupo, G., Alifano, P., Roberti, V., Bruni, C. B. & Bucci, C. Rab-interacting lysosomal protein (RILP): the Rab7 effector required for transport to lysosomes. *EMBO J.* **20,** 683–693 (2001).
14. Stenmark, H. *et al.* Inhibition of rab5 GTPase activity stimulates membrane fusion in endocytosis. *EMBO J.* **13,** 1287–1296 (1994).
15. Petiot, A., Faure, J., Stenmark, H. & Gruenberg, J. PI3P signaling regulates receptor sorting but not transport in the endosomal pathway. *J. Cell Biol.* **162,** 971–979 (2003).
16. Weir, M. L., Xie, H., Klip, A. & Trimble, W. S. VAP-A binds promiscuously to both v- and tSNAREs. *Biochem. Biophys. Res. Commun.* **286,** 616–621 (2001).
17. Pankiv, S. *et al.* FYCO1 is a Rab7 effector that binds to LC3 and PI3P to mediate microtubule plus end-directed vesicle transport. *J. Cell Biol.* **188,** 253–269 (2010).
18. Verhey, K. J., Kaul, N. & Soppina, V. Kinesin assembly and movement in cells. *Annu. Rev. Biophys.* **40,** 267–288 (2011).
19. Matsuzaki, F., Shirane, M., Matsumoto, M. & Nakayama, K. I. Protrudin serves as an adaptor molecule that connects KIF5 and its cargoes in vesicular transport during process formation. *Mol. Biol. Cell* **22,** 4602–4620 (2011).
20. Gil, J. E. *et al.* Phosphoinositides differentially regulate protrudin localization through the FYVE domain. *J. Biol. Chem.* **287,** 41268–41276 (2012).
21. Black, M. M. & Greene, L. A. Changes in the colchicine susceptibility of microtubules associated with neurite outgrowth: studies with nerve growth factor-responsive PC12 pheochromocytoma cells. *J. Cell Biol.* **95,** 379–386 (1982).
22. Zhang, C. *et al.* Role of spastin and protrudin in neurite outgrowth. *J. Cell. Biochem.* **113,** 2296–2307 (2012).
23. Arantes, R. M. & Andrews, N. W. A role for synaptotagmin VII-regulated exocytosis of lysosomes in neurite outgrowth from primary sympathetic neurons. *J. Neurosci.* **26,** 4630–4637 (2006).
24. Giordano, F. *et al.* $PI(4,5)P_2$-dependent and Ca^{2+}-regulated ER-PM interactions mediated by the extended synaptotagmins. *Cell* **153,** 1494–1509 (2013).
25. Mesmin, B. *et al.* A four-step cycle driven by PI(4)P hydrolysis directs sterol/PI(4)P exchange by the ER-Golgi tether OSBP. *Cell* **155,** 830–843 (2013).
26. Stefan, C. J. *et al.* Osh proteins regulate phosphoinositide metabolism at ER-plasma membrane contact sites. *Cell* **144,** 389–401 (2011).
27. Di Paolo, G. & De Camilli, P. Phosphoinositides in cell regulation and membrane dynamics. *Nature* **443,** 651–657 (2006).
28. Stenmark, H. Rab GTPases as coordinators of vesicle traffic. *Nature Rev. Mol. Cell Biol.* **10,** 513–525 (2009).
29. Zerial, M. & McBride, H. Rab proteins as membrane organizers. *Nature Rev. Mol. Cell Biol.* **2,** 107–117 (2001).
30. Pfeffer, S. R. Rab GTPase regulation of membrane identity. *Curr. Opin. Cell Biol.* **25,** 414–419 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Sagona and K.-W. Tan for assistance with plasmid constructs, E. Rønning for yeast two-hybrid analyses and protein purifications, Y. Zhen for advice on RAB7 knockdowns, and A. Engen for expert help with cell cultures. We are grateful to W. Do Heo for providing mCitrine–protrudin and the protrudin(FYVE4A) mutant. The Core Facilities for Advanced Light Microscopy and Electron Microscopy at Oslo University Hospital are acknowledged for providing access to relevant microscopes. C.R. and E.M.W. are senior research fellows of the Norwegian Cancer Society and South-Eastern Norway Regional Health Authority, respectively. C.B. was supported by the Associazione Italiana per la Ricerca sul Cancro (Investigator Grant 14709), Telethon-Italy (grant GGP09145) and MIUR (PRIN2010-2011). T.J. was supported by grant 196898 from the Norwegian Research Council and grant 71043-PR-2006-0320 from the Norwegian Cancer Society. H.S. was supported by grants from the Norwegian Cancer Society and an Advanced Grant from the European Research Council. This work was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 179571.

Author Contributions C.R. designed the project and the experiments, performed most of the high-content image analyses and quantitations, mapped the interaction surfaces of protrudin with RAB7(Q67L), made most of the figures and participated in writing the manuscript. E.M.W. performed all live imaging, SIM and LE–PM fusion experiments, and participated in CLEM and immunofluorescense imaging. N.M.P. performed all GFP-trap experiments and messenger RNA analyses. H.O. and T.J. made the RAB7 and FYCO plasmids and performed the MBP pull-down assays. S.W.S. performed the CLEM. M.V. performed some of the quantitations of protrusions in RPE1 cells. K.O.S. wrote scripts for image quantitations using Fiji or ImageJ, made plasmid constructs and stable cell lines and participated in quantitations. V.N. and C.B. performed GST pull-down experiments with protrudin and RAB proteins. A.B. performed conventional electron microscopy and electron microscopy tomography. H.S. coordinated the project and wrote the manuscript. All co-authors gave comments on the manuscript and approved the final version.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.R. (camilla.raiborg@rr-research.no) or H.S. (stenmark@ulrik.uio.no).

METHODS

Immunostaining and antibodies. Cells grown on coverslips were fixed in 3% paraformaldehyde for 15 min and washed twice in PBS containing 0.05% saponin. The cells were then stained using the indicated primary antibodies for 1 h, washed three times in PBS/saponin, stained with secondary antibodies for 1 h, and washed three times in PBS. The cells were mounted in Mowiol containing 2 μg ml^{-1} Hoechst 33342 (Sigma-Aldrich). Mouse anti-CD63 (H5C6) and mouse anti-LAMP1 (H4A3) were from the Developmental Studies Hybridoma Bank. Mouse anti-Myc (9E10) was from ATCC, rabbit anti-KIF5B (EPR10277(B)), rabbit anti-GFP (ab6556) and rabbit anti-Myc (ab9106) were from Abcam, mouse anti-GFP (11814 460001) was from Roche, mouse anti-α-tubulin (T9026), mouse anti-MBP (2H9) and rabbit anti-FYCO1 (HPA035526) were from Sigma. Rabbit anti-protrudin (Santa Cruz (H-21) or PTG (12680-1-AP)), goat anti-VAP-A (K-15), rabbit anti-KLC1 (H-75) and rabbit anti-RAB7 (H-50) were from Santa Cruz. Anti-FYCO1 mouse monoclonal antibody (A01) was from Abnova. Monoclonal anti-poly-His (H1029) was from Sigma-Aldrich. All secondary antibodies used for confocal studies were obtained from Jacksons ImmunoResearch Laboratories and secondary antibodies used for western blotting where obtained from LI-COR Biosciences GmbH and detected using the Odyssey developer system LI-COR Biosciences.

Plasmids and transfection. Plasmids used in this study (see Supplementary Table 1) were constructed by standard molecular biology methods. FYCO1 and RAB7 plasmids were described previously[17]. Point mutants were made using the QuickChange site-directed mutagenesis kit (Agilent Technologies). mCitrine–protrudin and protrudin(FYVE4A) were a gift from W. Do Heo[20]. pmTurquoise2–tubulin (Addgene 36202) and pmTurquoise–ER (Addgene 36204) were a gift from J. Goedhart[31]. Protrudin plasmids were generated by PCR using the cDNA clone BC030621 from ImaGene encoding the canonical human isoform 1 of protrudin (411 amino acids). PET-RAB5 (ref. 32) and PET-Rab7 (ref. 33) for bacterial expression of poly-His-tagged RAB GTPases were described previously. Plasmids were verified by sequencing and restriction digestion. siRNA-resistant protrudin was made resistant to siRNA #1 by introducing four point mutations by site-directed mutagenesis using quick-change PCR. FuGENE6 (HeLa) or FUGENE HD (hTERT-RPE-1, PC12) used for transient transfections were from Roche Diagnostics GmbH. Lipofectamine 3000 used for transient transfection of hTERT-RPE-1 cells was from Invitrogen.

Cell culture. All cell lines were from ATCC, recently authenticated and tested for mycoplasma contamination and grown according to ATCC guidelines. HeLa cells were grown on coverslips or in Nunc Lab-Tek II Chambered Coverglass or in MatTek 35 mm glass-bottom dishes, in DMEM containing 10% fetal calf serum (FCS), 2 mM glutamine, 100 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin in a 5% CO_2 incubator at 37 °C. hTERT-RPE-1 cells were grown in DMEM:F12, 10% (FCS) and 0.01 mg ml^{-1} hygromycin B. PC12 cells were grown in F-12K medium supplemented with 15% horse serum and 2.5% FCS. All media and reagents for cell culture were purchased from Gibco.

Generation of stable cell lines. HeLa and hTERT-RPE1 cells stably expressing GFP–FYCO1 were generated by lentiviral transduction. GFP–FYCO1 was subcloned into a Gateway ENTRY vector by standard molecular biology techniques. From this vector, a lentiviral transfer vector was generated by recombination with pCDH-EF1a-GW-IRES-Puro, a gateway-enabled derivative of pCDH-EF1a-MCS-IRES-Puro (Systems Biosciences). Lentivirus particles were packaged using a third-generation packaging system[34] as described (Addgene plasmid numbers 12251, 12253 and 12259)[35]. Cells were then transduced with low virus titres (multiplicity of infection (m.o.i.) < 1) and stable cell pools were generated by puromycin selection (0.5 μg ml^{-1} for HelaP, 5 μg ml^{-1} for hTERT-RPE1).

siRNA transfections. siRNA oligonucleotides were from Invitrogen Life Technologies or Quiagen (RAB7 and control for RAB7). All siRNA transfections were performed using RNAiMax (Invitrogen) according to the manufacturer's protocol and 20 or 50 nM siRNA oligonucleotide per well in HeLa or hTERT-RPE-1 cells, 100 nM in PC12 cells. Scrambled RNA used as negative controls has been described previously[36]. siRNA human FYCO1 was published before[17]. siRNA of SYT7 was published before[37]. For siRNA of human protrudin, the following sequences were used, modified from rat[7] to human. siRNA protrudin #1: 5′-CUUCUUGAUCCA GCUGGAGG[dT][dT]-3′; siRNA protrudin #2: 5′-CUGCCUGAUUCCGAGC UGAU[dT][dT]-3′; siRNA protrudin #3: 5′-CAGGUGGCAGAUGCCUUUGU GUU[dT][dT]-3′; siRNA human KLC1: 5′-AAAGAGCCCUCGAGAUCUA[U][U]-3′; siRNA human RAB7: 5′-CACGTAGGCCTTCAACACAAT-3′. For siRNA of rat protrudin, sequence #5 from[7] was used. For siRNA of rat FYCO1, the human sequence described previously[17] was translated to be rat specific: 5′-TCATCAAG AGCCAGTTCCT-3′. For siRNA of rat RAB7, 5′-CACGTAGGCCTTCAACAC GCG-3′. Combined transfections with siRNA and plasmid DNA were done sequentially with DNA added 3–18 h after siRNA depending on the experiment. For rescue of LAMP1 positioning in protrudin-depleted cells, after 6 h of siRNA transfection, HeLa cells were transiently transfected with low levels of siRNA-resistant Myc–protrudin for an additional 34 h. For quantification of protrusions in FYCO1-depleted cells, after 18 h of siRNA transfection, hTERT-RPE-1 cells were transfected with GFP–protrudin for an additional 36 h. For quantification of protrusions in RAB7-depleted cells, after 3 h of siRNA transfection, hTERT-RPE-1 cells were transfected with GFP–protrudin for an additional 30 h. For quantification of protrusions in SYT7-depleted cells, after 6 h of siRNA transfection, hTERT-RPE-1 cells were transfected with GFP–protruidin for an additional 42 h.

Quantitative real-time PCR of mRNA expression. mRNA expression analysis was done as described[38]. The primers used in the experiment were QT00195601 for SYT7 and QT00000721 for TBP (Qiagen).

Microscopy

High-content microscopy. An Olympus ScanR imaging platform with an UPLSAPO ×20 or a ×40 objective was used for quantitation of a large number of cells from PFA-fixed samples for the following assays:

Recruitment of protrudin to RAB7-positive endosomes in HeLa cells. Wide-field images of cells cotransfected with mCherry–RAB7(Q67L) and various GFP–protrudin plasmids were analysed automatically. When protrudin is coexpressed with RAB7(Q67L), protrudin is redistributed from a reticular ER localization pattern to cluster around RAB7(Q67L)-positive endosomes. The clustered appearance of protrudin was segmented automatically by ScanR software using edge-based detection, and the number of large protrudin punctae per cell was used as a measure of protrudin recruitment to RAB7(Q67L)-positive endosomes.

Relative peripheral localization of FYCO1 in HeLa cells. Sum projections of wide-field z-stack images of cells stably expressing GFP–FYCO1 or cells cotransfected with mCherry–FYCO1 and GFP alone or different GFP–protrudin plasmids were analysed automatically by ScanR software using intensity-based detection of mCherry–FYCO1 vesicles. The ratio between peripherally localized FYCO1 and perinuclear FYCO1 (as segmented by relative distance to Hoechest positive nuclei, as defined by edge-based detection) was calculated by the total intensity of segmented mCherry–FYCO1 vesicles in the indicated regions of GFP-positive cells.

LAMP1 positioning in siRNA-transfected HeLa and RPE1 cells. Perinuclearly localized clusters of LAMP1-positive vesicles were segmented automatically by ScanR software using intensity-based detection. Individual LAMP1 vesicles were not segmented. The total number of cells was quantified automatically using edge-based detection of Hoechst-positive nuclei and used to calculate the percentage of cells with perinuclear clusters of LAMP1. For rescue experiments, cells transfected with low levels of siRNA-resistant Myc–protrudin were picked automatically by the analysis software based on intensity of the stained Myc tag in the rescue construct, and the level of perinuclear LAMP1 clusters was directly compared to siRNA-treated, Myc–protrudin-negative neighbouring cells within each image.

Protrusions in hTERT-RPE-1 cells. Maximum projections of wide-field z-stack images were inspected manually and cells with protrusions were identified by the signal from GFP–protrudin, which was overexposed to see the thin protrusions. The total number of transfected cells were analysed automatically by ScanR software based on GFP intensity, and used to calculate the percentage of transfected cells exhibiting protrusions.

Confocal fluorescence microscopy. Cells were grown on coverslips, fixed in 3% PFA and immunostained as described earlier. The cells were examined with a Zeiss LSM710 or LSM780 confocal microscope (Carl Zeiss MicroImaging GmbH) equipped with an Ar-Laser Multiline (458/488/514 nm), a DPSS-561 10 (561 nm), a Laser diode 405-30 CW (405 nm), and a HeNe-laser (633 nm). The objective used was a Zeiss plan-Apochromat ×63/1.4 Oil DIC III. Image processing and visualization were performed with ZEN 2010 (Carl Zeiss MicroImaging GmbH), Photoshop CS4 (Adobe), ImageJ (National Institutes of Health) and Imaris 7.1.1 (Bitplane AG).

Super-resolution microscopy. HeLa cells were transfected with GFP–protrudin and mCherry–FYCO1. Twenty-four hours later, the cells were fixed with PFA and stained against KIF5B (rabbit-anti-KIF5B, Abcam) and GFP (mouse-anti-GFP, Roche) and embedded in ProLongGold Antifade Reagent (Life Technologies). We performed three-dimensional SIM imaging on an OMX V4 system (DeltaVision OMX Microscope Applied Precision, GE Healthcare) equipped with an Olympus ×60 NA 1.42 objective and three cooled sCMOS cameras. Using the 488, 568 and 642 nm laser lines to excite the used fluorophores, cells were illuminated with a grid pattern and for each image plane, 15 images were collected (five phases for three rotations of the illumination pattern). z-Stacks were acquired with a z-spacing of 125 nm. Super-resolution images were reconstructed from the raw image files, aligned and projected using Softworx software (Applied Precision, GE Healthcare) and processed in ImageJ for presentation.

Live-cell imaging. Long-term live-cell imaging experiments to observe changes in cell morphology were done with RPE-1 cells on a Delta Vision Deconvolution microscope (Applied Precision, GE Healthcare) using a ×40 objective and 2 × 2 binning. Cells were maintained at 37 °C and 5% CO_2 and stacks comprising the whole cell were acquired. Images were deconvolved and projected using Softworx software (Applied Precision, GE Healthcare) and processed in ImageJ for presentation. Live-cell imaging to monitor vesicle movement in HeLa cells was done similarly,

but a $\times$60 objective without binning was used. Images were acquired at 0.5 Hz for 2 min. Vesicles were tracked using the ImageJ plugin 'Manual tracking'. The tracks were further processed with the Chemotaxis and Migration Tool 2.0 from Ibidi GmbH.

Triple-fluorescence live-cell imaging of mCitrine–protrudin, mCherry–FYCO1 and mTq2–tubulin or mTq2–ER was done on an OMX V4 system (DeltaVision OMX Microscope Applied Precision, GE Healthcare) equipped with an Olympus $\times$60 NA 1.42 objective and three cooled sCMOS cameras. mTurquoise2-, mCitrine- and mCherry-tagged proteins were excited in conventional mode with an InsightSSI light source (Applied Precision) and suitable filter sets. z-Stacks with a spacing of 200 nm were acquired with a frame rate of 0.33 Hz. Deconvolution and alignment was performed using Softworx software (Applied Precision, GE Healthcare).

Live-cell imaging of GFP–protrudin and mCherry–FYCO1 was done similarly, but with a frame rate of 1 Hz. Images were acquired simultaneously in both channels using the multiple cameras of the OMX system to avoid imaging artefacts that could potentially be introduced by fast moving vesicles and sequential image acquisition.

Electron microscopy. Cells for electron tomography experiments were grown on coverslips and transfected with Myc–protrudin and RAB7(Q67L) as described. Fixation was done with 2% glutaraldehyde in 0.1 M PHEM (240 mM PIPES, 100 mM HEPES, 8 mM $MgCl_2$, 40 mM EGTA, pH 6.9). After postfixation with 1% OsO_4 and uranylacetate en bloc staining, semithin sections (150–170 nm) were cut parallell to the substrate and placed on carbon/formvar-coated slotgrids (EMS). Samples were observed in a FEI Tecnai microscope at 120 kV and image series taken between -58° and 58° with 2° increment. Series were recorded around two orthogonal axes with a FEI camera at 1.9 nm pixel size. Single-axis tomograms were computed using weighted back projection and merging into a dual-axis tomogram was performed using the IMOD package[39]. Display, segmentation and animation of tomograms were also performed using IMOD software. Final video processing was done in iMovie11 (version 9.0.9, Apple).

For CLEM, HeLa cells transiently transfected with GFP–protrudin and mCherry–RAB7(Q67L) were grown on photo-etched coverslips (Electron Microscopy Sciences). Cells were fixed in 4% formaldehyde, 0.1% glutaraldehyde/0.1 M PHEM, for 1 h. The coverslips were washed with 0.1 M PHEM buffer and mounted with Mowiol containing 2 μg ml^{-1} Hoechst 33342. The cells were examined with a Zeiss LSM780 confocal microscope (Carl Zeiss MicroImaging GmbH) using an Ar-Laser Multiline (458/488/514 nm), a DPSS-561 10 (561 nm), a Laser diode 405-30 CW (405 nm). Cells of interest were identified by fluorescence microscopy, a 4.9-μm-thick z-stack (voxel size: 51 nm $\times$ 51 nm $\times$ 140 nm) was acquired and the optical data were deconvolved using the Huygens Essential deconvolution software (64 bit, version 4.5, Scientific Volume Imaging B.V.). The relative positioning of the cells on the photo-etched coverslips was determined by taking a DIC image. The coverslips were removed from the object glass, washed with 0.1 M PHEM buffer and fixed in 2% glutaraldehyde/0.1 M PHEM for 1 h. Cells were postfixed in osmium tetroxide, stained with tannic acid, dehydrated stepwise to 100% ethanol and flat-embedded in Epon. Serial sections ($\sim$100 nm) were cut on an Ultracut UCT ultramicrotome (Leica), collected on formvar-coated slot-grids, and poststained with lead citrate.

Sections were observed at 80 kV in a JEOL-JEM 1230 electron microscope and images were recorded using iTEM software with a Morada camera (Olympus). Overlapping images taken at $\times$12,000 magnification were stitched together using the TrakEM2 plugin for ImageJ.

Analysis of KIF5B intensity in FYCO1 vesicles by fluorescence microscopy. The analysis was done on confocal images by manually making a circle (region of interest (ROI)) enclosing mCherry-positive vesicles. For comparison, the mean intensity of KIF5B was divided by the mean intensity of mCherry–FYCO1 at every ROI to adjust for FYCO1 levels.

Western blotting. Cells were rinsed twice with ice-cold PBS before being lysed in 25 mM HEPES (pH 7.2), 125 mM potassium acetate, 2.5 mM magnesium acetate, 5 mM EGTA, 0.5% NP40, 1 mM dithiothreitol (DTT) and protease inhibitor cocktail. Proteins were separated by SDS–PAGE on TGX Precast gels (Biorad) and blotted with PVDF membranes from Merck Millipore. Immunodetection was performed using fluorescently labelled secondary antibodies and Odyssey developer.

Assay of interaction between protrudin LCR and RAB7. Bacterially purified His–RAB7 or His–RAB5 (5 μg) loaded with GTP or GDP was incubated with GST or GST–protrudin in PBS with 2 mM $MgCl_2$ and 0.8 mM GTP or GDP for 2 h at 4 °C on a rotating wheel. Subsequently, samples were subject to GST pull-down using a glutathione resin and then subjected to SDS–PAGE and western blotting using anti-GST and anti-His-tag antibodies.

GFP-trap assay. GFP trap (GFP-trap magnetic beads, ChromoTek) was used for interaction studies and the experiments were performed according to the manufacturer's protocol. Stable cell lines (HeLa or RPE1) expressing GFP–FYCO1 (Fig. 3d–f) or transiently transfected HeLa cells (Fig. 3b, c and Extended Data Figs 1d, e, 4c) were used. Owing to space limitations, inputs were omitted from some of the GFP traps. GFP–FYCO1 in the trap serves as an internal loading control. One per cent input was used in most of the GFP traps unless otherwise indicated.

MBP pull-down assays. HEK293 cells (1.8×10^5) in 6-cm plates were transfected with 1 μg of each plasmid using Metafectene PRO (Biontex). After 24 h, cells were rinsed twice in PBS and lysed in modified RIPA buffer (50 mM Tris-Cl pH 7.4, 150 mM NaCl, 1% NP40, 0.25% DOC, 2 mM EDTA) added protease inhibitor cocktail (Complete Mini-EDTA-free; Roche) and phosphatase inhibitor cocktail set II (Calbiochem). Extracts were centrifuged at 13,000 r.p.m., 10 min at 4 °C and precleared using 10 μl empty amylose beads before incubation on a rotating wheel for 1.5 h at 4 °C with recombinant MBP–FYCO1$^{585-1233}$ bound to amylose beads. Following five times washing with lysis buffer, western blots were performed and developed using anti-GFP (ab290; Abcam) or anti-MBP (M6295 ; Sigma) antibodies. MBP pull-down experiments of *in vitro* translated proteins were performed as recently described[40].

Assay of interaction between protrudin LCR and RAB7. Bacterially purified His–RAB7 or His–RAB5 loaded with GTP or GDP was incubated with GST or GST–protrudin in PBS with 2 mM $MgCl_2$ and 0.8 mM GTP or GDP for 2 h at 4 °C on a rotating wheel. Subsequently, samples were subject to GST pull-down using a glutathione resin and then subjected to SDS–PAGE and western blotting using anti-GST and anti-His-tag antibodies (Extended Data Fig. 1f) and Coomassie and anti-RAB7 (Fig. 1d).

Analysis of neurite outgrowth in PC12 cells. PC12 cells grown in F-12K medium supplemented with 15% horse serum and 2.5% FCS were seeded in Nunc Lab-Tek II Chambered Coverglass and transfected with plasmid DNA 3 days after seeding, using FUGENE HD. Eighteen hours after transfection, the medium was changed to DMEM/F12 supplemented with 1% horse serum and 50 ng ml^{-1} NGF for 30 h. For siRNA experiments, PC12 cells were transfected twice with 100 nM siRNA oligonucleotide using RNAiMax before stimulation for 48 h with 100 ng ml^{-1} NGF. The cells were fixed in 3% PFA, stained with 2 μg ml^{-1} Hoechst 33342, anti-tubulin and rhodamine-Phalloidin and mounted in Ibidi mounting medium (Ibidi, GmbH). Cells were imaged on a Delta Vision Deconvolution microscope (Applied Precision, GE Healthcare) equipped with Elite TruLight Illumination System, a CoolSNAP HQ2camera and a 20× Plan Apochromat lens. Images (29 z-sections 0.7 μm apart) were acquired and deconvolved using the softWoRx software (Applied Precision). Deconvolved projections were analysed by visual inspection using Fiji. For PC12 cells transfected with plasmid DNA, cells with neurite outgrowth were defined as cells where protrusions were extended beyond a circle of 30 μm in diameter from the cell centre. To this end, an ImageJ macro was used that generated a ROI with a diameter of 30 μm around the manually selected centre of the cell. Protrusions extending beyond the boundaries of this ROI were scored. Imaris was used to make a surface view of a confocal z-stack of a PC12 cell expressing mCherry–FYCO1 and GFP–protrudin. Small background dots, not related to the cell, were removed for clarity. For siRNA-treated PC12 cells stained for actin and tubulin, the actual length of neurites was measured using Fiji.

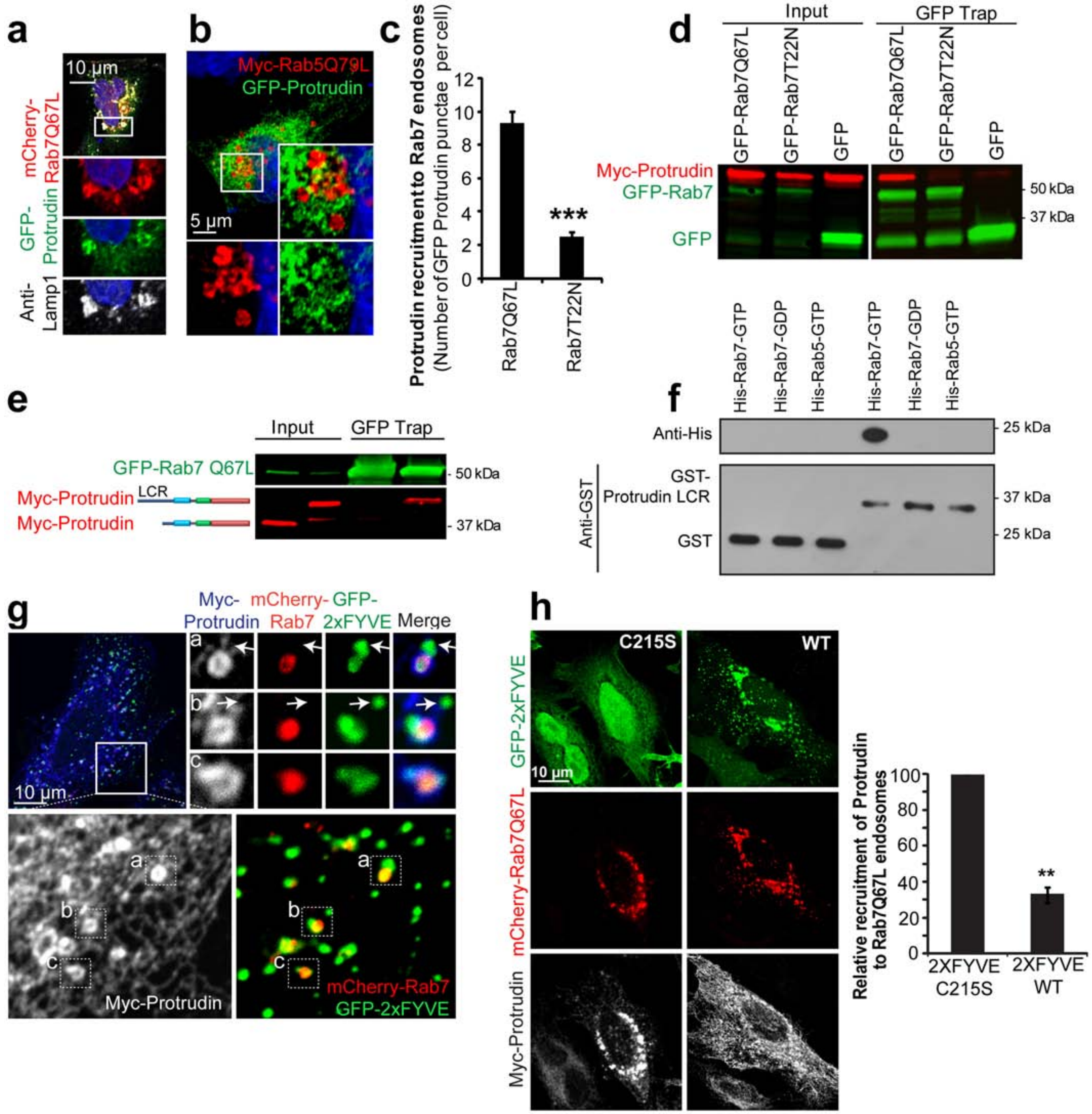
Analysis of vesicle velocity and intensities from live cells. A customized Fiji script written in Jython was used to measure the intensity and speed of FYCO1-positive LEs over time from live cell videos of GFP–protrudin and mCherry–FYCO1. Vesicles were first tracked manually using the « Point ROI »-Function of Fiji. This resulted in a list of *xyt* coordinates corresponding to the vesicle localization over time. Using this list, the mean intensity of the individual channels was automatically measured in a ROI of 0.4 μm diameter around each tracked spot. Vesicle velocity was calculated from the distance the vesicle moved between time points and the known time resolution of the video.

Analysis of LE fusion with the plasma membrane. To measure the fusion of LE with the plasma membrane, the lysosomal transmembrane protein LAMP1 was fused to the GFP-derivative pHluorin on its luminal (amino-terminal) domain and to mCherry on its cytoplasmic (carboxy-terminal) domain. pHluorin was thereby inserted between the signal peptide and the transmembrane domain of LAMP1. HeLa cells were cotransfected with pHluorin–LAMP1–mCherry, Cerulean–FYCO1, mCitrine–protrudin or mCitrine–protrudin(FYVE4A), or empty plasmids as indicated. For knockdown experiments, siRNAs were transfected 4 h before complementary DNA transfections. Forty-eight hours after transfection, cells were fixed with 3% PFA and stained without permeabilization with a mouse-anti-GFP antibody recognizing the surface-exposed pHluorin and a secondary antibody against mouse coupled to Al647. Fluorescence images of cells co-expressing all three constructs were acquired and total fluorescence of the Al647 channel divided by total fluorescence in the mCherry channel was used to quantify the amount of surface-localized LAMP1 normalized to total expression of the pHluorin–LAMP1–mCherry construct. At least 20 cells per condition were analysed and the experiments were repeated 3 times.

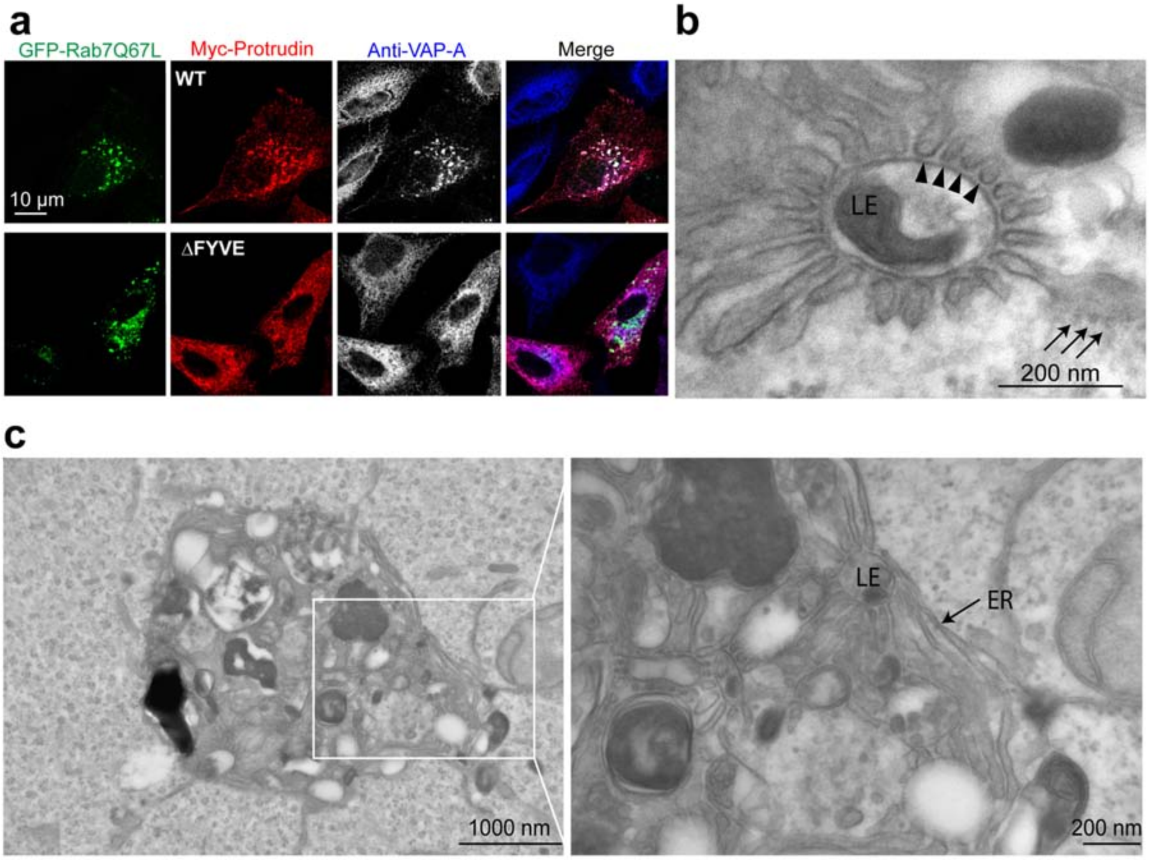
Statistical analysis and considerations. The number of individual experiments and the number of cells analysed are indicated. The number of experiments was adapted to the expected effect size and the anticipated consistency between experiments.

One-sample t-test was used to analyse the level of significance when the control was set to 1 or 100. Unpaired t-test was used to test two samples with equal variance. For the calculation of the correlation between vesicle speed and protrudin intensity, Spearman's rank correlation coefficient was used. All error bars denote mean values ± s.e.m. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

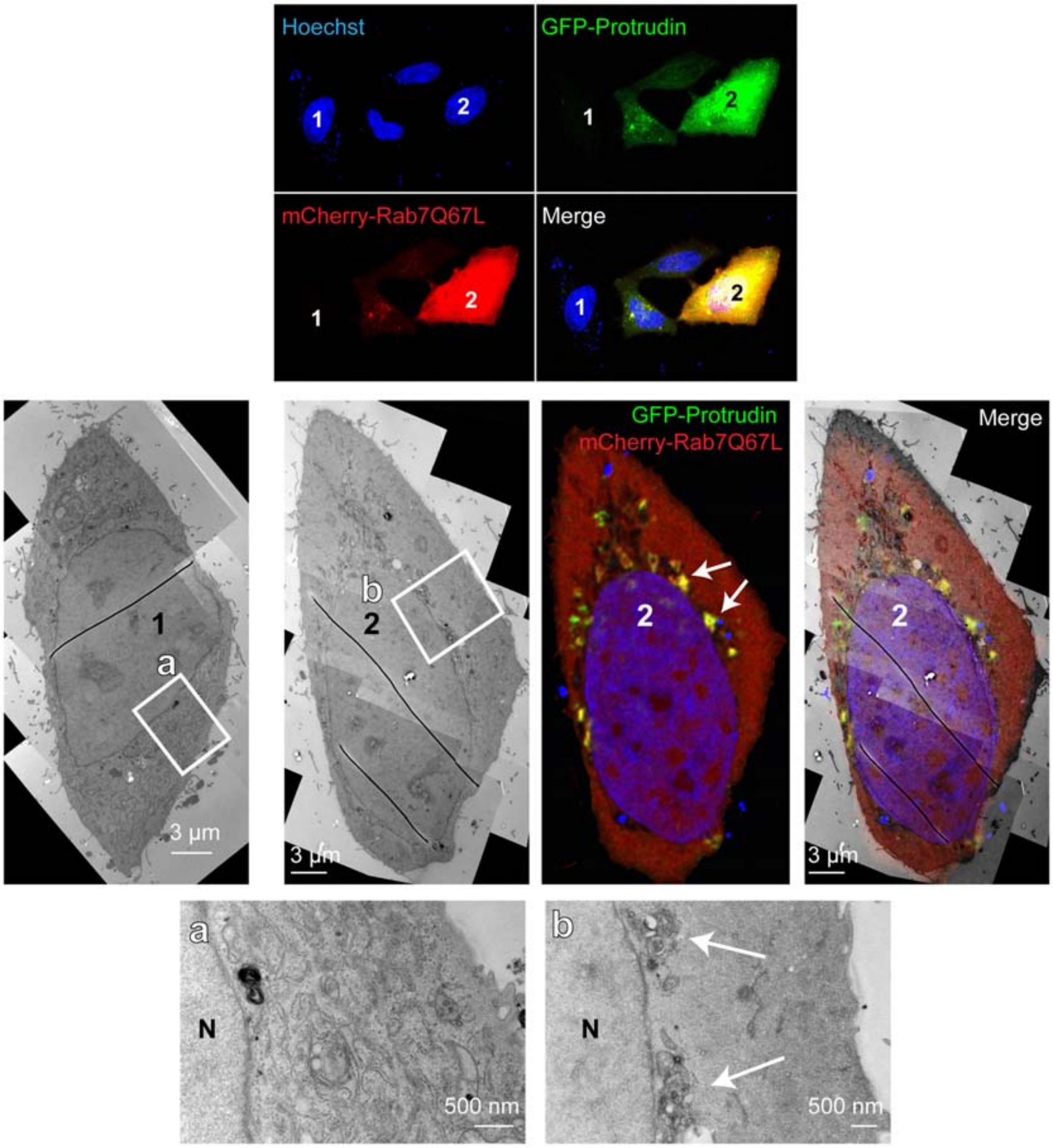
31. Goedhart, J. *et al.* Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nature Commun.* **3,** 751 (2012).
32. Chiariello, M., Bruni, C. B. & Bucci, C. The small GTPases Rab5a, Rab5b and Rab5c are differentially phosphorylated *in vitro*. *FEBS Lett.* **453,** 20–24 (1999).
33. Spinosa, M. R. *et al.* Functional characterization of Rab7 mutant proteins associated with Charcot-Marie-Tooth type 2B disease. *J. Neurosci.* **28,** 1640–1648 (2008).
34. Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72,** 8463–8471 (1998).
35. Campeau, E. *et al.* A versatile viral system for expression and depletion of proteins in mammalian cells. *PLoS ONE* **4,** e6529 (2009).
36. Cabezas, A., Bache, K. G., Brech, A. & Stenmark, H. Alix regulates cortical actin and the spatial distribution of endosomes. *J. Cell Sci.* **118,** 2625–2635 (2005).
37. Colvin, R. A. *et al.* Synaptotagmin-mediated vesicle fusion regulates cell migration. *Nature Immunol.* **11,** 495–502 (2010).
38. Pedersen, N. M. *et al.* The PtdIns3P-binding protein phafin 2 mediates epidermal growth factor receptor degradation by promoting endosome fusion. *Traffic* **13,** 1547–1563 (2012).
39. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116,** 71–76 (1996).
40. Alemu, E. A. *et al.* ATG8 family proteins act as scaffolds for assembly of the ULK complex: sequence requirements for LC3-interacting region (LIR) motifs. *J. Biol. Chem.* **287,** 39275–39290 (2012).



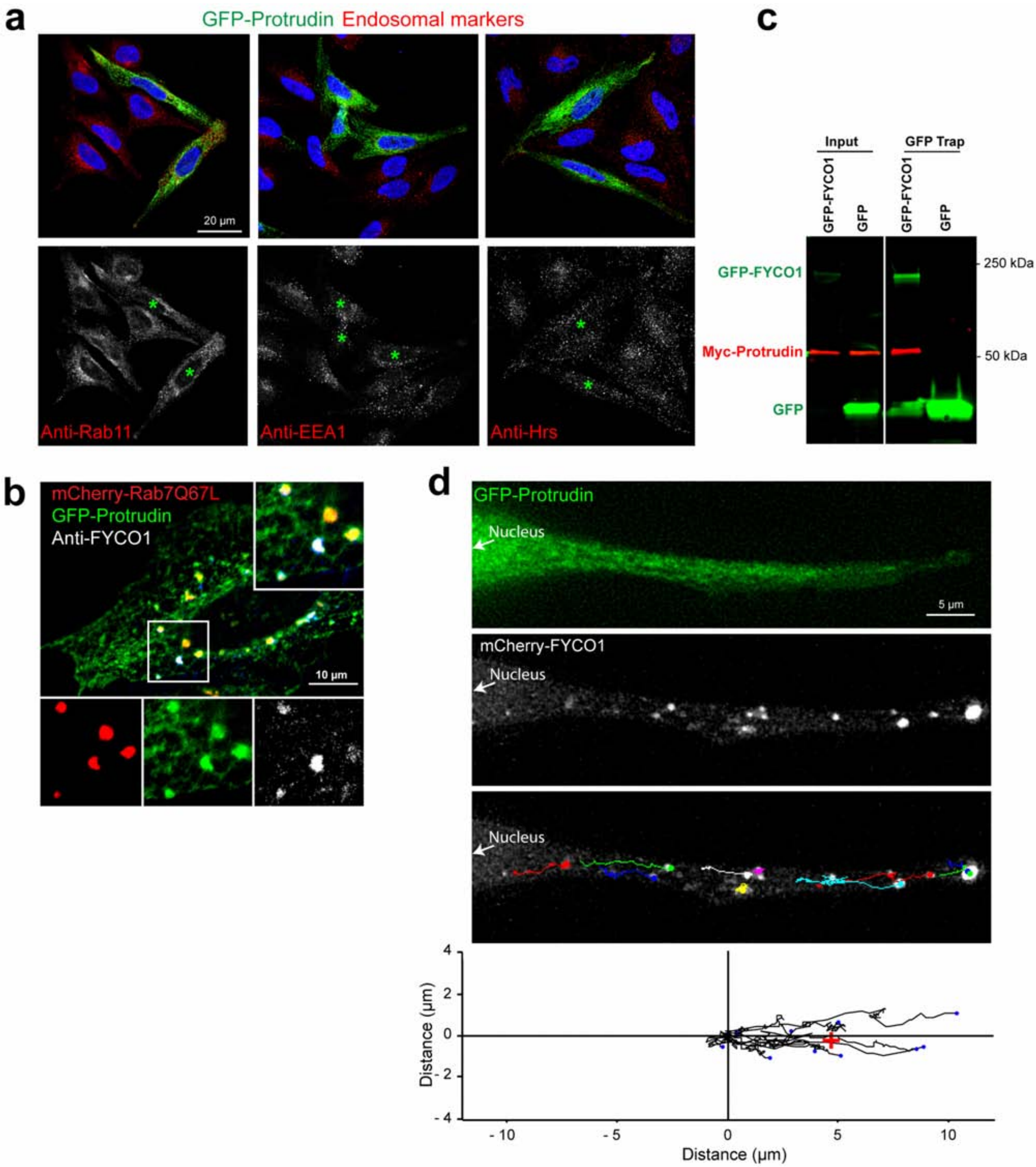
Extended Data Figure 1 | Protrudin forms RAB7- and PtdIns(3)P-dependent ER–LE contact sites. a, RAB7(Q67L) causes excessive recruitment of GFP–protrudin to LAMP1-positive endosomes in HeLa cells. Confocal images are representative of at least 10 captures. **b**, GFP–protrudin is not recruited to RAB5(Q79L)-positive endosomes. Confocal images are representative of at least 10 captures. **c**, Quantification of GFP–protrudin recruitment to RAB7 endosomes in cells cotransfected with mCherry–RAB7(Q67L) or RAB7(T22N). Quantification of Olympus ScanR images. Error bars denote ± s.e.m. RAB7(Q67L): 4 experiments, 2,673 cells; RAB7(T22N): 3 experiments, 1,814 cells. $***P < 0.001$ (unpaired *t*-test). **d**, Protrudin is pulled down with GFP–RAB7(Q67L), but not with GFP–RAB7(T22N) or GFP alone in a GFP-trap assay. **e**, The C-terminal part of Myc–protrudin, including LCR domains, is pulled down with GFP–RAB7(Q67L) in a GFP-trap assay. **f**, *In vitro* pulldown of a GST–protrudin–LCR construct (residues 208–274) shows direct binding to His–RAB7 loaded with GTP, but not to RAB7–GDP or RAB5–GTP. **g**, Protrudin-positive ER associates with RAB7/PtdIns(3)P-positive endosomes in HeLa cells transfected with low levels of Myc–protrudin, mCherry–RAB7 and GFP–2×FYVE, shown by confocal microscopy. Note that PtdIns(3)P-positive RAB7-negative endosomes are protrudin-negative (arrows). Confocal images are representative of at least 10 captures. **h**, Confocal micrographs showing that high levels of 2×FYVE compete with the recruitment of protrudin to RAB7(Q67L)-positive endosomes. Representative of 4 confocal captures and 240 wide-field images per condition. Quantification of Olympus ScanR images. Error bars denote ± s.e.m. from 3 experiments. GFP–2×FYVE^{C215S}: 5,861 cells, GFP–2×FYVE: 4,641 cells. $**P < 0.01$ (one-sample *t*-test).



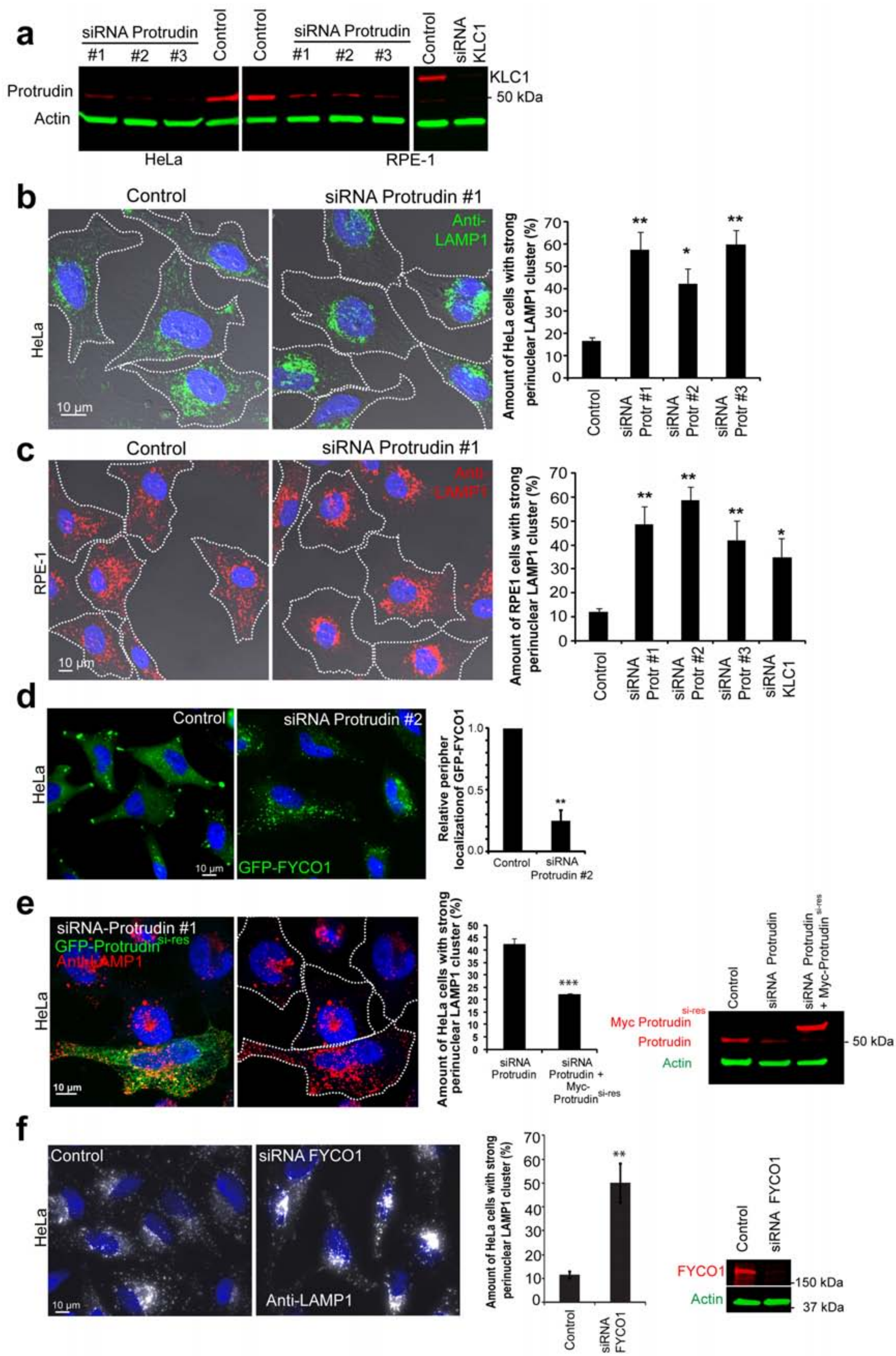
Extended Data Figure 2 | LEs and ER form extensive contact sites upon expression of RAB7(Q67L). **a**, Protrudin co-recruits the ER marker VAP-A to RAB7(Q67L) endosomes in a FYVE-dependent manner. HeLa cells were transfected with GFP–RAB7(Q67L) and Myc–protrudin or FYVE-deleted protrudin as shown by confocal microscopy. Images are representative of at least 10 captures. WT, wild type. **b**, **c**, Electron micrographs showing that coexpression of protrudin and RAB7(Q67L) leads to formation of endosomal clusters with strong recruitment of ER structures. The massive accumulation of ER–endosome contact sites on an endosome can be clearly seen (arrowheads in **b**). Continuity with the rough ER is indicated (arrows in **b**). Representative of 30 images.



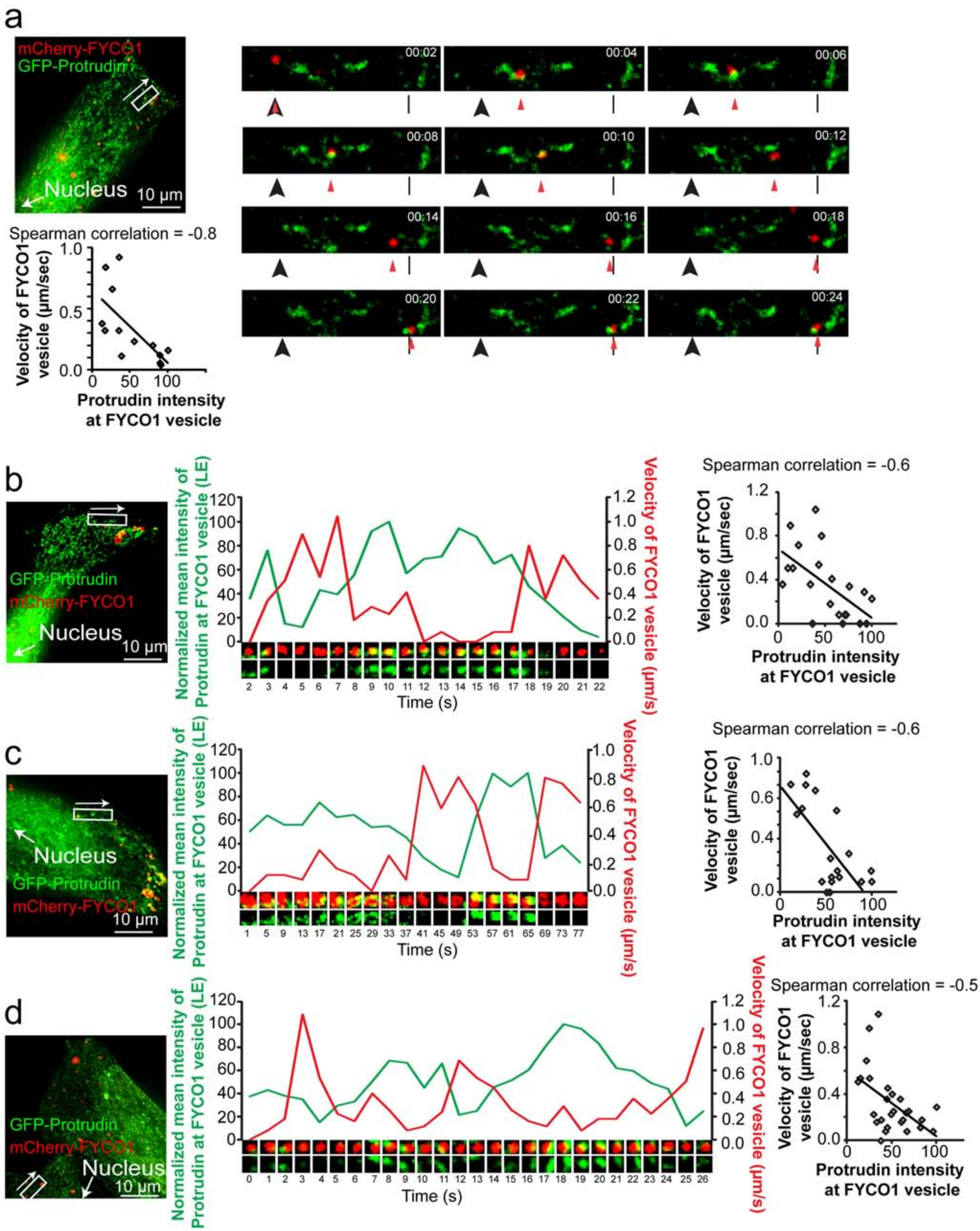
Extended Data Figure 3 | Redistribution of ER membranes in protrudin- and RAB7(Q67L)-transfected cells. CLEM. HeLa cells were transfected with GFP–protrudin and mCherry–RAB7(Q67L), and four neighbouring cells with varying levels of overexpression were analysed by confocal microscopy before electron microscopy. A non-transfected cell (cell number 1, inset a) shows ER membranes distributed all over the cytosol. A cell expressing high levels of protrudin and RAB7(Q67L) (cell number 2, inset b) shows a complete redistribution of ER membranes to perinuclear structures. These structures are positive for protrudin and RAB7(Q67L) (arrows) as seen by the immunofluorescense-electron microscopy overlay and represent the structures seen at higher resolution in Fig. 1e and Extended Data Fig. 2b, c). Representative of 4 CLEM images.



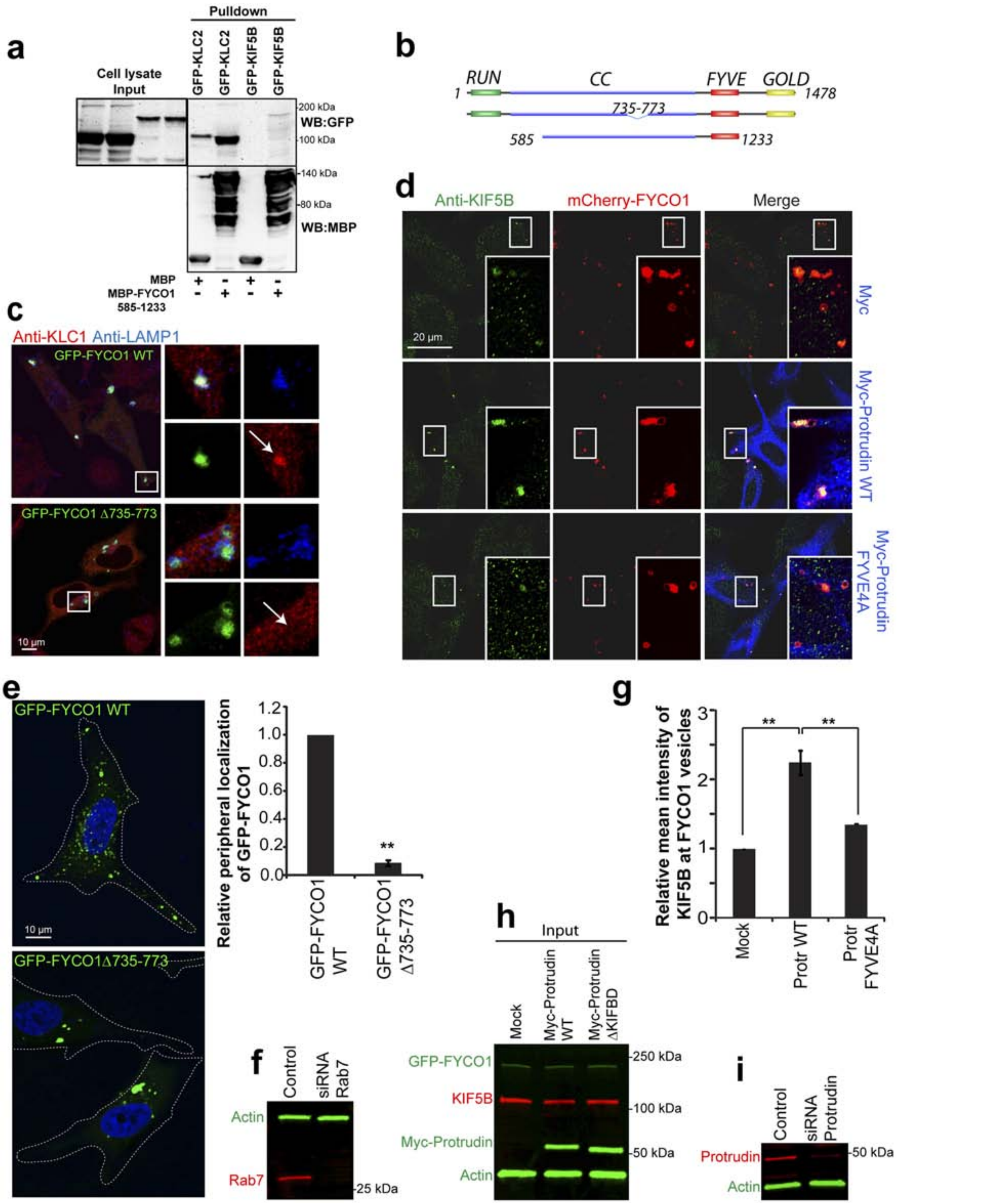
Extended Data Figure 4 | Protrudin mediates LE translocation to the cell periphery. **a,** The positioning of early and recycling endosomes is not changed in GFP–protrudin-overexpressing HeLa cells (indicated by asterisks). Confocal images are representative of at least 10 captures. **b,** Endogenous FYCO1 localizes to protrudin and RAB7(Q67L)-positive endosomes. HeLa cells were transfected with GFP–protrudin and mCherry–RAB7(Q67L), stained with anti-FYCO1 antibody and imaged by confocal microscopy. Images are representative of at least 10 captures. **c,** Myc–protrudin is pulled down with GFP–FYCO1 but not with GFP in a GFP-trap assay. Input 1%. **d,** Still pictures from Supplementary Video 2, representative of 10 videos. A distal part of a GFP–protrudin- and mCherry–FYCO1-positive HeLa cell is shown. FYCO1 vesicles move towards the microtubule plus-end of the cell. The movement of each vesicle during 120 s is shown by colour-coded lines. The coloured dots indicate where the vesicles stop. Bottom panel: example for the calculation of the centre of mass of all tracked vesicles from Supplementary Video 2. The starting point of every track is set to $x = 0$ and $y = 0$. The endpoints of all tracks (blue dots) define the centre of mass (red cross) as their spatial averaged point.


a
siRNA Protrudin
#1 #2 #3
Control
Control
siRNA Protrudin
#1 #2 #3
Control
siRNA KLC1
Protrudin
Actin
KLC1
50 kDa
HeLa
RPE-1
b
Control
siRNA Protrudin #1
Anti-LAMP1
HeLa
10 µm
Amount of HeLa cells with strong perinuclear LAMP1 cluster (%)
Control
siRNA Protr #1
siRNA Protr #2
siRNA Protr #3
c
Control
siRNA Protrudin #1
Anti-LAMP1
RPE-1
10 µm
Amount of RPE1 cells with strong perinuclear LAMP1 cluster (%)
Control
siRNA Protr #1
siRNA Protr #2
siRNA Protr #3
siRNA KLC1
d
Control
siRNA Protrudin #2
HeLa
10 µm
GFP-FYCO1
Relative peripher localizationof GFP-FYCO1
Control
siRNA Protrudin #2
e
siRNA-Protrudin #1
GFP-Protrudin si-res
Anti-LAMP1
HeLa
10 µm
Amount of HeLa cells with strong perinuclear LAMP1 cluster (%)
siRNA Protrudin
siRNA Protrudin + Myc-Protrudin si-res
Control
siRNA Protrudin
siRNA Protrudin + Myc-Protrudin si-res
Myc Protrudin si-res
Protrudin
Actin
50 kDa
f
Control
siRNA FYCO1
HeLa
10 µm
Anti-LAMP1
Amount of HeLa cells with strong perinuclear LAMP1 cluster (%)
Control
siRNA FYCO1
Control
siRNA FYCO1
FYCO1
Actin
150 kDa
37 kDa

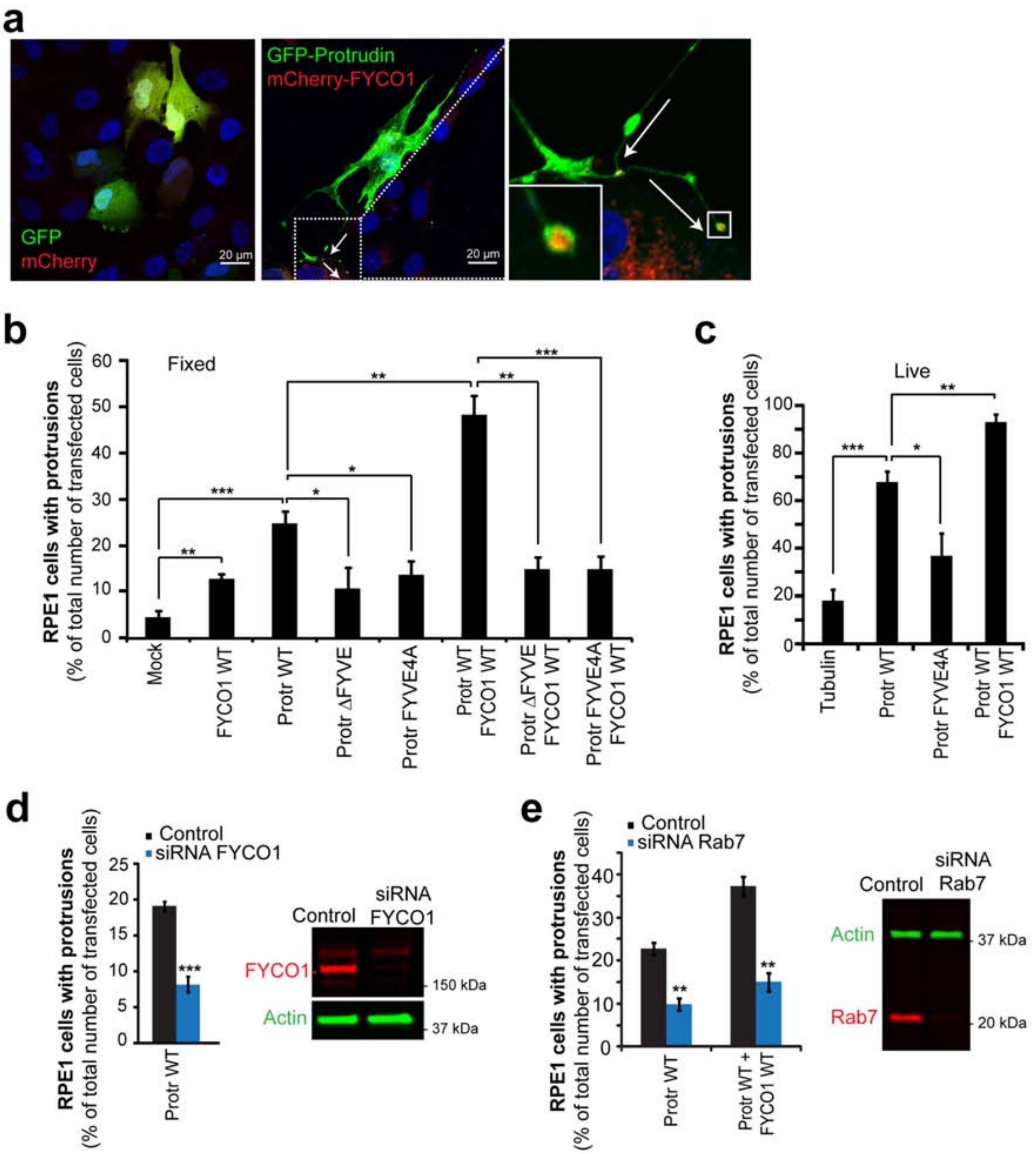
Extended Data Figure 5 | LAMP1 and FYCO1 cluster perinuclearly in cells depleted of protrudin. **a**, Western blot showing the level of siRNA-mediated knockdown of protrudin and KLC1 as compared to control. **b**, **c**, HeLa (**b**) and RPE1 (**c**) cells depleted of protrudin by siRNA show perinuclear localization of LAMP1-positive LEs. Images are representative of at least 4 confocal and 290 wide-field captures. Cells were transfected with control siRNA, siRNA targeting protrudin (#1–3) or KLC1 (only RPE1) and analysed by Olympus ScanR high-throughput microscopy. Error bars denote ± s.e.m. HeLa, 3 experiments: control: 5,945 cells; siRNA protrudin #1: 2,965 cells; siRNA protrudin #2: 3,378 cells; siRNA protrudin #3: 3,986 cells. RPE1, 4 experiments: control: 3,655 cells; siRNA protrudin #1: 2,292 cells; siRNA protrudin #2: 1,961 cells; siRNA protrudin #3: 2,402 cells; siRNA KLC1 (3 experiments): 1,625 cells. *$P < 0.05$, **$P < 0.01$ (unpaired t-test). **d**, GFP–FYCO1 clusters perinuclearly in protrudin siRNA cells as compared to control cells, in which GFP–FYCO1 is mainly localized to the cell periphery as quantified by Olympus ScanR microscopy (representative of 240 images). Error bar denotes ± s.e.m. from 3 experiments. Control: 2,979 cells; siRNA protrudin: 2,011 cells. **$P < 0.01$ (one-sample t-test). **e**, In protrudin siRNA-treated cells, late endosomes cluster perinuclearly, but not in neighbouring cells expressing siRNA-resistant Myc–protrudin, as shown by confocal microscopy (representative of 2 confocal and 290 wide-field captures). The graph shows quantification of perinuclear localization of LAMP1 in siRNA-treated cells (siRNA protrudin #1) compared to neighbouring cells expressing siRNA-resistant Myc–protrudin by Olympus ScanR analysis. Error bars denote ± s.e.m. from 3 experiments. siRNA protrudin: 3,601 cells; siRNA protrudin/Myc–protrudin: 1,437 cells. ***$P < 0.001$ (unpaired t-test). Knockdown efficiency and level of overexpression of the siRNA-resistant Myc–protrudin construct is shown by western blotting. **f**, LAMP1-positive LEs cluster perinuclearly in cells depleted of FYCO1 as shown by Olympus ScanR quantitative microscopy (representative of 290 images). Error bars denote ± s.e.m. from 3 experiments. Control: 5,033 cells; siRNA FYCO1: 5,820 cells. **$P < 0.01$ (unpaired t-test). Knockdown efficiency of FYCO1 is shown by western blotting.



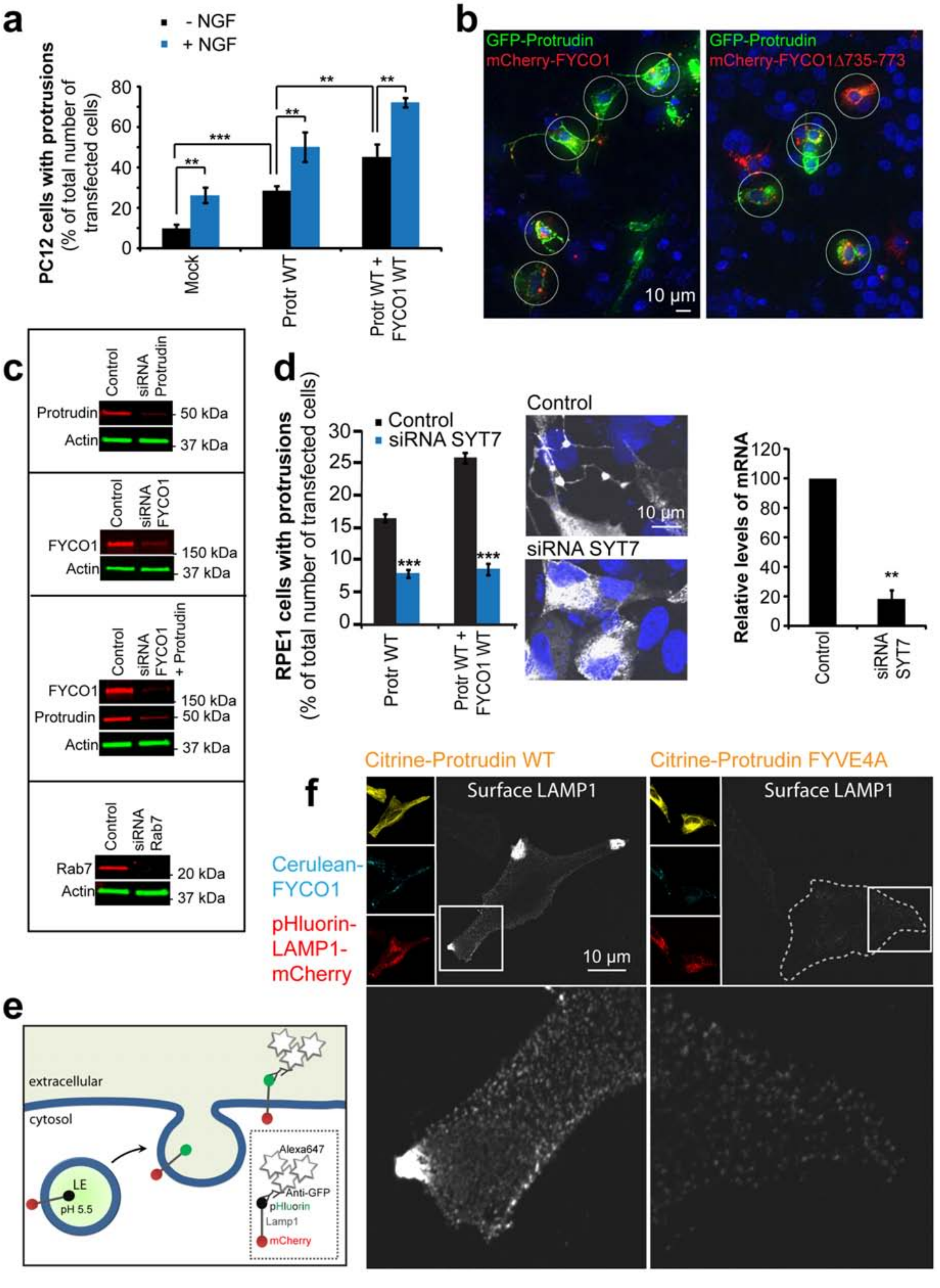
Extended Data Figure 6 | Protrudin forms repeated contact sites with LEs and mediates their translocation to the cell periphery. a–d, Still images from videos showing how mCherry–FYCO1 vesicles move towards the plasma membrane while dynamically associating with GFP–protrudin (representative of 31 videos). The vesicles halt when close to protrudin, speed up when escaping protrudin and reassociate with a new pool of protrudin as they slow down. Images are captured every second. Owing to space limitations, selected frames are shown in **a** and **c**. The scatter plots show the inverse correlation of speed versus protrudin association of the FYCO1 vesicles, as indicated by a linear regression line.



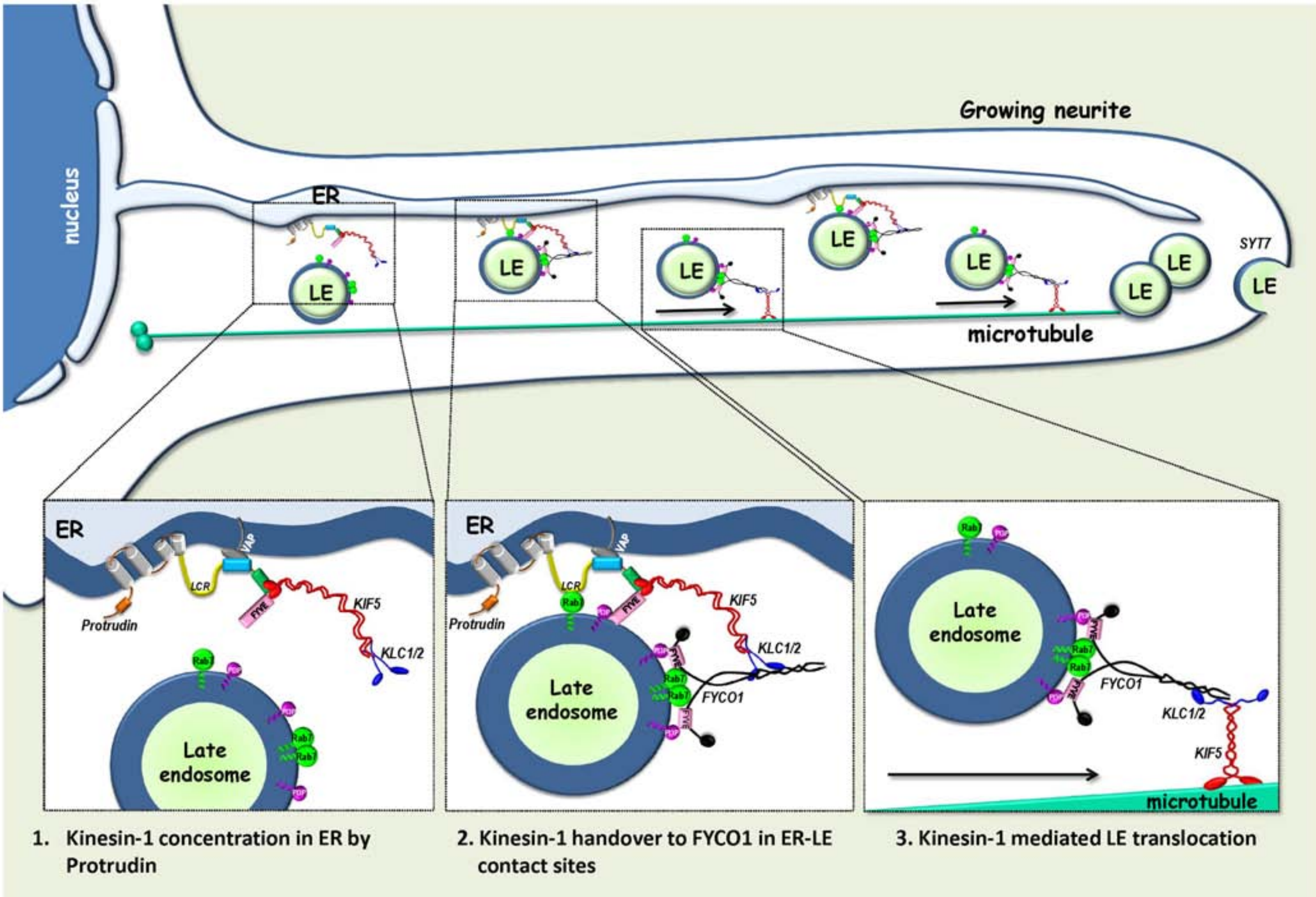
Extended Data Figure 7 | Protrudin loads kinesin 1 onto FYCO1 on LEs. **a**, MBP–FYCO1$^{585–1233}$ beads capture GFP–KLC2 or GFP-KIF5B from cell extracts. Input: 10%. **b**, Domain structure of FYCO1 plasmids used in this study. **c**, Confocal images showing the colocalization of KLC1 and LAMP1 with GFP–FYCO1, which is lost in the GFP–FYCO1(Δ735–773) mutant. Confocal images are representative of at least 10 captures. **d**, Endogenous kinesin 1 colocalizes with mCherry–FYCO1-positive vesicles and accumulates on such vesicles in cells coexpressing Myc–protrudin. HeLa cells were transfected with mCherry–FYCO1 and different Myc–protrudin constructs and analysed by confocal microscopy. Images are representative of at least 30 captures. Quantified in Extended Data Fig. 7g. **e**, GFP–FYCO1(Δ735–773)-positive LEs are unable to translocate to the cell periphery. Representative of 5 confocal and 440 wide-field images. Quantification was done using an Olympus ScanR microscope. Error bar denotes ± s.e.m. from 3 experiments. GFP–FYCO1: 1,150 cells; GFP–FYCO1(Δ735–773): 1,372 cells. $^{**}P < 0.01$ (one-sample *t*-test). **f**, Western blot showing siRNA-mediated downregulation of RAB7 in HeLa cells stably expressing GFP–FYCO1. **g**, Quantification of the relative mean intensity of endogenous KIF5B on mCherry–FYCO1 vesicles from confocal images in Extended Data Fig. 7d. Error bars denote ± s.e.m. Mock: 5 experiments (133 cells, 353 vesicles) were set to 1; Myc–protrudin: 5 experiments (118 cells, 217 vesicles); Myc–protrudin(FYVE4A): 3 experiments (90 cells, 193 vesicles). $^{**}P < 0.01$. Mock/protrudin, one-sample *t*-test; protudin/FYVE4A, unpaired *t*-test. **h**, Western blot showing 1% of the input used for the GFP-trap assay from HeLa cells stably expressing GFP–FYCO1 (Fig. 3e). **i**, Western blot showing siRNA-mediated downregulation of protrudin in RPE1 cells stably expressing GFP–FYCO1.



Extended Data Figure 8 | Protrudin and FYCO1 cooperatively induce protrusion formation. **a**, Overexpression of GFP–protrudin induces a protrusion phenotype in RPE1 cells (see also Supplementary Video 6) and co-expression of GFP–protrudin and mCherry–FYCO1 increases the protrusion phenotype. mCherry–FYCO1 vesicles localize to the very end of each protrusion. Arrows indicate the direction of two protrusions pointing at the FYCO1-positive ends. Cells were transfected with the indicated constructs for 48 h and analysed by confocal microscopy. Images are representative of at least 15 confocal captures. Quantified in Extended Data Fig. 8b. **b**, Quantification of the number (%) of cells with protrusions in RPE1 cells transfected with the indicated constructs for 48 h. Images were acquired with an Olympus ScanR microscope and cells with protrusions were quantified by visual inspection of GFP. Error bars denote ± s.e.m. from 4 experiments. GFP + mCherry: 1,131 cells; GFP + mCherrry–FYCO1: 1,203 cells; GFP–protrudin + mCherry: 1,226 cells; GFP–protrudin(ΔFYVE) + mCherry: 871 cells; GFP–protrudin(FYVE4A) + mCherry: 1,365 cells; GFP–protrudin + mCherry–FYCO1: 979 cells; GFP–protrudin(ΔFYVE) + mCherry–FYCO1: 766 cells; protrudin(FYVE4A) + mCherry–FYCO1: 1,365 cells. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ (unpaired *t*-test). **c**, Quantification of the number of cells with protrusions in RPE1 cells transfected with the indicated constructs and measured by live imaging (example in Supplementary Video 6). Cells with protrusions were quantified by visual inspection of GFP in the videos. Error bars denote ± s.e.m. Tubulin: 3 experiments, 19 cells; GFP–protrudin: 4 experiments, 33 cells; GFP–protrudin(FYVE4A): 3 experiments, 24 cells; GFP–protrudin + mCherry–FYCO1: 2 experiments, 31 cells; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ (unpaired *t*-test). **d**, Knockdown of FYCO1 reduces the number of cells forming protrusions upon overexpression of protrudin. FYCO1 knockdown efficiency is shown by western blotting. Error bars denote ± s.e.m. from 5 experiments. Control: 1,706 cells; siRNA FYCO1: 1,393 cells. $***P < 0.001$ (unpaired *t*-test). **e**, Protrudin-mediated protrusion formation is dependent on RAB7. Quantification of the number (%) of cells with protrusions in RPE1 cells transfected with the indicated constructs. Images were acquired with an Olympus ScanR microscope, and cells with protrusions were quantified by visual inspection. RAB7 knockdown efficiency is shown by western blotting. Error bars denote ± s.e.m. from 3 experiments. Control GFP–protrudin + mCherry: 2,631 cells; siRNA RAB7 GFP–protrudin + mCherry: 2,428 cells; control GFP–protrudin + mCherry–FYCO1: 3,046 cells; siRNA RAB7 GFP–protrudin + mCherry–FYCO1: 3,207 cells. $**P < 0.01$ (unpaired *t*-test).


a
- NGF
+ NGF
PC12 cells with protrusions (% of total number of transfected cells)
Mock
Protr WT
Protr WT + FYCO1 WT
b
GFP-Protrudin
mCherry-FYCO1
GFP-Protrudin
mCherry-FYCO1Δ735-773
10 μm
c
Control
siRNA Protrudin
Protrudin
Actin
50 kDa
37 kDa
siRNA FYCO1
FYCO1
150 kDa
siRNA FYCO1 + Protrudin
siRNA Rab7
Rab7
20 kDa
d
RPE1 cells with protrusions (% of total number of transfected cells)
Control
siRNA SYT7
Protr WT
Protr WT + FYCO1 WT
Relative levels of mRNA
e
extracellular
cytosol
LE
pH 5.5
Alexa647
Anti-GFP
pHluorin
Lamp1
mCherry
f
Citrine-Protrudin WT
Citrine-Protrudin FYVE4A
Surface LAMP1
Cerulean-FYCO1
pHluorin-LAMP1-mCherry

Extended Data Figure 9 | Protrudin and FYCO1 cooperatively induce neurite outgrowth by facilitating LE fusion with the plasma membrane. **a**, Protrudin overexpression induces neurite outgrowth in PC12 cells. NGF treatment and overexpression of FYCO1 increase protrudin-mediated neurite outgrowth. WT, wild type. Error bars denote ± s.e.m. from 3 experiments. Without NGF: GFP + mCherry: 336 cells; GFP–protrudin + mCherry: 326 cells; GFP–protrudin + mCherry–FYCO1: 215 cells. With NGF: GFP + mCherry: 333 cells; GFP–protrudin + mCherry: 224 cells; GFP–Protrudin + mCherry–FYCO1: 173 cells. $^{**}P < 0.01$, $^{***}P < 0.001$ (unpaired *t*-test). **b**, Deconvolved projections of z-stacks recorded by a DeltaVision microscope showing PC12 cells with protrusions (mediated by protrudin and FYCO1, left panel representative of 80 images) and without protrusions (inhibited by mutant FYCO1(Δ735–773), right panel representative of 30 images). Whereas mCherry–FYCO1 is transported into forming protrusions, mCherry–FYCO1(Δ735–773) (unable to engage kinesin 1) is not. The circles have a diameter of 30 μm. Cells with protrusions reaching out of the circle are regarded as protrusion-positive. Quantified in Fig. 4c. **c**, Western blots showing levels of knockdown in PC12 cells. **d**, Knockdown of SYT7 reduces the number of RPE1 cells with protrusions upon overexpression of protrudin or coexpression of protrudin and FYCO1. Images were acquired with an Olympus ScanR microscope, and cells with protrusions were quantified by visual inspection. Error bars denote ± s.e.m. from 3 experiments. Control GFP–protrudin + mCherry: 4,278 cells; siRNA SYT7 GFP–protrudin + mCherry: 901 cells; control GFP–protrudin + mCherry–FYCO1: 3,045 cells; siRNA SYT7 GFP–protrudin + mCherry–FYCO1: 1,731 cells. $^{***}P < 0.001$ (unpaired *t*-test). Confocal images are shown (representative of at least 10 confocal and 240 wide-field captures). Knockdown efficiency of SYT7 is shown by quantitative realtime PCR. Error bar denotes ± s.e.m. from 3 experiments. $^{**}P < 0.01$ (one-sample *t*-test). **e**, Surface-accumulated LAMP1 is detected with a reporter construct consisting of a double-tagged LAMP1 construct. The surface-exposed GFP variant pHluorin is recognized with an anti-GFP-antibody on non-permeabilized cells. Indirect immunofluorescence amplifies the surface signal. **f**, Confocal images (representative of at least 30 captures) of cells expressing Cerulean–FYCO1 and the reporter construct pHluorin–LAMP1–mCherry show accumulation of surface LAMP1 in the presence of mCitrine–protrudin wild type, but not mCitrine–protrudin(FYVE4A). For quantification, the total fluorescence in the Alexa647 channel was normalized to the total fluorescence in the mCherry channel to take expression levels of the reporter construct into account. Quantified in Fig. 4e, f.



Extended Data Figure 10 | Model for the functions of protrudin and FYCO1 in LE translocation and neurite outgrowth. The kinesin-1-binding ER protein protrudin contacts LEs via coincident detection of RAB7–GTP and PtdIns(3)P on LEs. This allows kinesin 1 to be handed over to the RAB7–GTP- and PtdIns(3)P-binding LE protein FYCO1, which in turn mediates the transport of LEs towards the plasma membrane along microtubules. LEs are reloaded with kinesin 1 at ER–LE contact sites as the vesicles move into the growing neurite and ultimately fuse with the plasma membrane by a SYT7-dependent mechanism.

doi:10.1038/nature14122

EZH2 inhibition sensitizes *BRG1* and *EGFR* mutant lung tumours to TopoII inhibitors

Christine M. Fillmore[1,2,3], Chunxiao Xu[4,5], Pooja T. Desai[1], Joanne M. Berry[1], Samuel P. Rowbotham[1,2,3], Yi-Jang Lin[2], Haikuo Zhang[4,5], Victor E. Marquez[6], Peter S. Hammerman[4], Kwok-Kin Wong[4,5] & Carla F. Kim[1,2,3]

Non-small-cell lung cancer is the leading cause of cancer-related death worldwide[1]. Chemotherapies such as the topoisomerase II (TopoII) inhibitor etoposide effectively reduce disease in a minority of patients with this cancer[2,3]; therefore, alternative drug targets, including epigenetic enzymes, are under consideration for therapeutic intervention[4]. A promising potential epigenetic target is the methyltransferase EZH2, which in the context of the polycomb repressive complex 2 (PRC2) is well known to tri-methylate histone H3 at lysine 27 (H3K27me3) and elicit gene silencing[5]. Here we demonstrate that EZH2 inhibition has differential effects on the TopoII inhibitor response of non-small-cell lung cancers *in vitro* and *in vivo*. *EGFR* and *BRG1* mutations are genetic biomarkers that predict enhanced sensitivity to TopoII inhibitor in response to EZH2 inhibition. *BRG1* loss-of-function mutant tumours respond to EZH2 inhibition with increased S phase, anaphase bridging, apoptosis and TopoII inhibitor sensitivity. Conversely, *EGFR* and *BRG1* wild-type tumours upregulate *BRG1* in response to EZH2 inhibition and ultimately become more resistant to TopoII inhibitor. *EGFR* gain-of-function mutant tumours are also sensitive to dual EZH2 inhibition and TopoII inhibitor, because of genetic antagonism between *EGFR* and BRG1. These findings suggest an opportunity for precision medicine in the genetically complex disease of non-small-cell lung cancer.

To validate that EZH2 is an important target for non-small-cell lung cancer (NSCLC), we generated a 116-gene lung cancer *EZH2* co-expression gene signature (Supplementary Table 1). This signature had predictive power for cancer progression using the Director's Challenge data set of 416 human lung adenocarcinomas[6], partly because of stratification of later-stage tumours to the *EZH2* high group (Extended Data Fig. 1a). To control for this covariate, exclusively stage 1 and moderately differentiated tumours were examined, confirming that the signature could robustly further stratify patients into risk groups (Fig. 1a). Gene ontology analysis revealed that the *EZH2* co-expression signature was highly enriched for cell cycle, DNA synthesis and DNA repair genes (Supplementary Table 2). One of the genes highly co-expressed with *EZH2* in primary tumours was topoisomerase 2A (*TOP2A*), which encodes the TopoII helicase targeted by etoposide.

To test EZH2 inhibition as a therapy for NSCLC, *EZH2* expression was stably knocked down with one of two different small hairpins in a panel of NSCLC cell lines. Western blot confirmed that EZH2 protein and catalytic mark, H3K27me3, were decreased in each transduced cell line and could be rescued by *EZH2* expression from a second lentivirus (Fig. 1b and Extended Data Fig. 1b). We then determined etoposide half-maximum inhibitory concentration (IC_{50}) at 4 days. Of the seven lines, HCC15, A549, H157 and PC9, termed 'sensitized' lines, had lower etoposide IC_{50} when *EZH2* was knocked down. Conversely, H460, H23 and Sw1573 cell lines, termed 'protected' lines, had higher etoposide IC_{50} as shEZH2 lines (Fig. 1c). Rescue of EZH2 levels completely abrogated the change in etoposide IC_{50} driven by the 3′ untranslated region (UTR) targeting hairpin (A549 and Sw1573; Fig. 1c, grey bars). The sensitized and protected phenotypes were not due to differential degrees of *EZH2* knockdown (Extended Data Fig. 1b, c).

Next, we used pharmacological EZH2 inhibition via the *S*-adenosylhomocystein hydrolase inhibitor, DZNep, which causes proteosomal degradation of PRC2 components including EZH2 (refs 7, 8) and the specific EZH2 methyltransferase inhibitor, GSK126 (ref. 9). Western blot confirmed that 4 days of 1 μM DZNep effectively reduced EZH2 protein and H3K27me3, and 10 μM GSK126 for 4 days or 2 μM GSK126 for 9 days caused a decrease in H3K27me3 levels yet EZH2 remained unchanged (Fig. 1d and Extended Data Fig. 2a). Fourteen of 26 NSCLC cell lines were more sensitive to 4-day etoposide in the presence of 1 μM DZNep, while the other lines were less sensitive to the chemotherapy in the presence of DZNep (Fig. 1e and Extended Data Fig. 2b). For the sensitized lines, pretreatment with 2 μM GSK126 for 9 days sensitized the lines to 4-day etoposide with continued GSK126 treatment (14 days total). For the protected lines, 10 μM of GSK126 for 4 days best recapitulated the etoposide protection caused by DZNep and shEZH2 (Fig. 1e and Extended Data Fig. 2c). IC_{50} shift results were validated with the Chou–Talalay combination index[10], demonstrating strong synergism (combination index < 0.48) between DZNep and etoposide as well as synergism (combination index < 0.64) between GSK126 and etoposide (Fig. 1f and Supplementary Table 3). The combination index assay also confirmed drug antagonism (combination index > 1) in the protected lines.

We examined the mutational annotation available for the NSCLC lines and found that 12 of 14 sensitized cell lines harboured inactivating mutations in *BRG1* (*SMARCA4*) or activating mutations in *EGFR*, while 10 of 12 protected cell lines were wild type (WT) for the two genes (Supplementary Table 4 and Extended Data Fig. 2c; Fisher's exact test, $P = 0.001$). Cell lines segregated into the same genotype-specific protected and sensitized classes when a different TopoII inhibitor, doxorubicin[11], was combined with DZNep (Extended Data Fig. 2d).

To determine whether the protected and sensitized phenotypes could be observed *in vivo*, we treated xenograft-bearing mice with etoposide and EZH2 inhibition (EZH2i). For the sensitized *BRG1* mutant cell line H157, early treatment with dual etoposide and DZNep therapy prevented tumours from forming in four out of six mice, proving more efficacious than etoposide or DZNep alone (Fig. 2a and Extended Data Fig. 3a, b). In contrast, the protected H23 xenografts that received early dual therapy grew significantly larger than those treated with either DZNep or etoposide alone (Fig. 2b and Extended Data Fig. 3b). Furthermore, in mice with established *EGFR*-driven PC9 xenografts, the combination of GSK126 and etoposide prevented tumour growth (Fig. 2c).

Next, mouse models of lung cancer predicted to be sensitized ($EGFR^{T790M;L858R}$ transgenic; *EGFR* hereafter[12]) or protected ($Kras^{G12D/+}$; $p53^{\Delta/\Delta}$; *Kras*/*p53* hereafter[13]) tumour types were treated with DZNep and etoposide. The *Kras*/*p53* model, WT for *Brg1* and *Egfr*, represents a predicted 'protected' cancer, whereas the *EGFR* model, driven by oncogenic *EGFR*, represents a predicted 'sensitized' cancer. Etoposide,

[1]Stem Cell Program, Boston Children's Hospital, Boston, Massachusetts 02115, USA. [2]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Harvard Stem Cell Institute, Cambridge, Massachusetts 02138, USA. [4]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [5]Belfer Institute for Applied Cancer Science, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [6]Chemical Biology Laboratory, National Cancer Institute, National Institutes of Health, Frederick, Maryland 21702, USA.

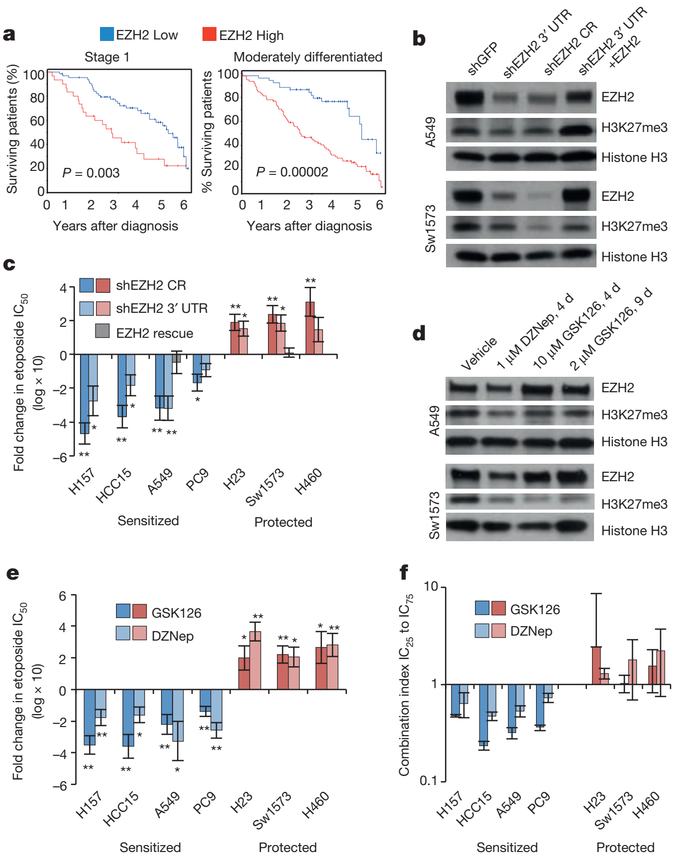


Figure 1 | EZH2i sensitizes BRG1 or EGFR mutants to TopoIIi. a, Director's Challenge samples were hierarchically clustered into two risk groups using the *EZH2* co-expression signature (Supplementary Table 1). The Kaplan–Meier curves only for stage 1 ($n = 94$) or only for moderately differentiated tumours ($n = 142$) to 6 years after diagnosis are shown. **b**, Western blot for EZH2 and H3K27me3 on indicated transduced lines; total histone H3 is shown as loading control. CR indicates a coding region targeting hairpin. **c**, Fold change ± s.e.m. in etoposide IC_{50} between transduced lines ($n = 3$ biological replicates for HCC15, A549, PC9, H23 and Sw1573, $n = 4$ biological replicates for HCC15 and H460, rescues $n = 3$ biological replicates; $*P < 0.04$, $**P < 0.01$). **d**, Western blot for EZH2 and H3K27me3 on lines treated with indicated drugs. **e**, Fold change ± s.e.m. in etoposide IC_{50} between vehicle-treated and drug-treated lines ($n = 3$ biological replicates for all except $n = 4$ biological replicates for H157 + DZNep; $*P < 0.04$, $**P < 0.01$). **f**, Average Chou–Talalay combination index values ± s.e.m. (also see Supplementary Table 3) for fractions affected equivalent to IC_{25}–IC_{75} ($n = 3$ biological replicates).

DZNep, or combination therapy was then administered to randomized cohorts of mice with radiographically documented lung masses for 4 weeks (Fig. 2d). Marked tumour regression in the *EGFR* model was observed in response to 4 weeks of dual etoposide and DZNep treatment, while mice in the other treatment arms showed continued tumour growth (Fig. 2e and Extended Data Fig. 4a). In striking contrast, the *Kras*/*p53* tumours proceeded to grow despite dual treatment (Fig. 2f). DZNep efficacy was confirmed by EZH2 immunohistochemistry for both models (Extended Data Fig. 4b, c).

To address the mechanism through which EZH2i changed sensitivity to TopoII inhibitor (TopoIIi), we considered the physical interaction between BRG1 and TopoII that allows for increased TopoII function[14]. Because BRG1 and EZH2 are known to be genetically antagonistic[15], we hypothesized that protected cell lines upregulated *BRG1* in response to EZH2i and thereby had increased TopoII function. *BRG1* transcript measured by quantitative PCR with reverse transcription (RT-qPCR) was reproducibly increased by DZNep treatment, although *BRG1* levels were not significantly different when the cells were treated with GSK126 (Fig. 3a). To assess the function of BRG1-containing BAF complexes we quantified anaphase bridges, which are known to indicate a failure of TopoII to decatenate DNA before mitosis and can be attributable to BAF complex dysfunction (Extended Data Fig. 5a). DZNep or GSK126 treatment increased anaphase bridges in *BRG1* mutant cell lines and decreased anaphase bridges in WT cells (Fig. 3b and Extended Data Fig. 5b). *EGFR* mutant cells, despite BRG1 upregulation, also had increased anaphase bridging with DZNep or GSK126 treatment and showed high levels of EGFR in dividing cells (Extended Data Fig. 5c).

We next examined cell cycle and apoptosis dynamics of the lines. While the protected lines showed no difference in apoptotic levels in etoposide compared with dual-treated cultures, the sensitized lines had significantly higher apoptotic fractions in dual-treated cultures than in cultures treated with etoposide as a single agent (Fig. 3c). Furthermore, protected etoposide-treated lines had an average of 13.5% fewer cells in

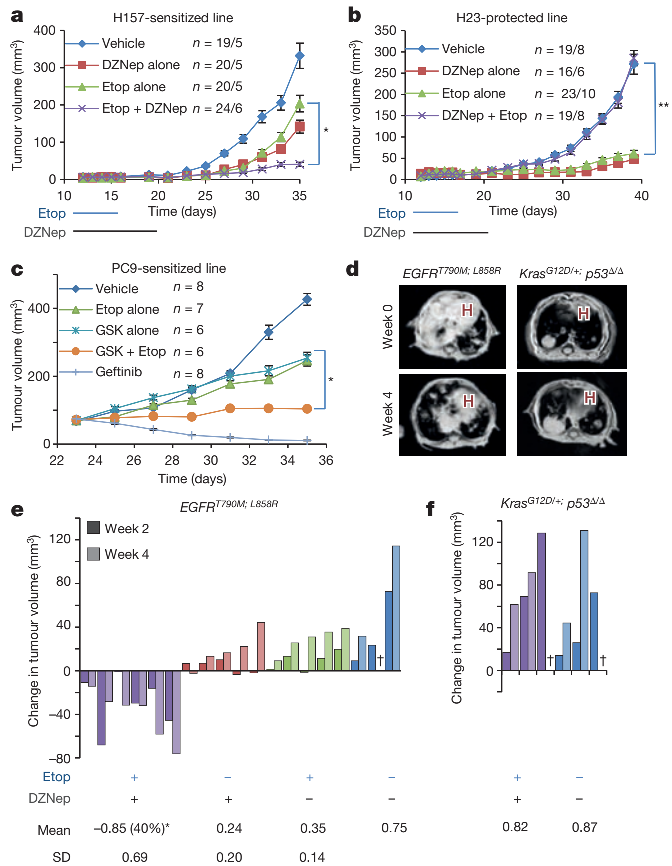


Figure 2 | *In vitro* sensitivities to EZH2i plus TopoIIi predict *in vivo* responses. Either the H157 (**a**) or the H23 (**b**) cell line was injected into the flanks of Nude mice and tumours were allowed to form. On day 12, mice were randomly segregated into cohorts that received either placebo, DZNep, etoposide or dual therapy for 2 weeks, and average tumour size ± s.e.m. were plotted (n for tumours/mice in each arm indicated on graphs; $*P = 0.002$, $**P = 0.0005$ dual versus etoposide). **c**, The PC9 cell line was injected into the flanks of Nude mice and tumours were allowed to grow to 70 mm^3. Mice were then treated with etoposide, GSK126, dual therapy or gefitinib (as a positive control) for 2 weeks, and average tumour size ± s.e.m. were plotted (n indicated on legend, mice with one tumour each; $P < 0.008$ for dual versus etoposide or GSK126 alone). **d**, Representative magnetic resonance images of mice of indicated genotypes on combination etoposide plus DZNep treatment at 0 and 4 weeks after treatment initiation. H, heart area. **e**, Waterfall plot depicting tumour growth ± s.e.m. of $EGFR^{T790M;L858R}$ tumours after 2 weeks and 4 weeks of treatment with vehicle (blue), etoposide (green), DZNep (red) and etoposide plus DZNep (purple). The y axis indicates percentage tumour growth versus day 0. Each bar represents an individual mouse (†the mouse died before the magnetic resonance imaging time point). Statistical analyses were performed on the 4-week log_2-transformed data ($P = 0.008$ dual versus DZNep and $P = 0.004$ dual versus etoposide). **f**, Waterfall plot depicting tumour growth ± s.e.m. of $Kras^{G12D/+}$;$p53^{\Delta/\Delta}$ tumours after 2 weeks and 4 weeks of treatment with vehicle (blue) and etoposide plus DZNep (purple).

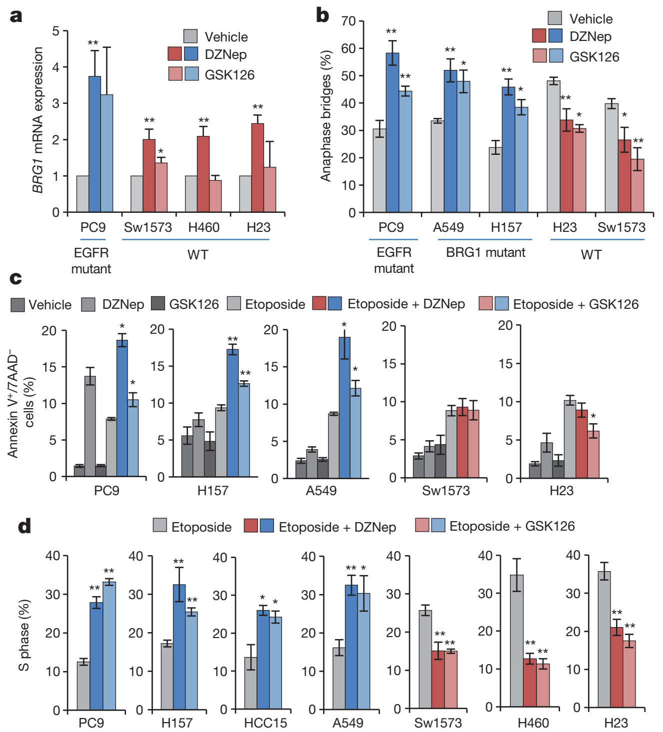


Figure 3 | Dual EZH2i and TopoIIi differentially affects cell cycle, apoptosis and anaphase bridging. a, RT-qPCR for average expression of *BRG1* ± s.e.m. in indicated cell lines in response to 4 days of 1 μM DZNep or 10 μM GSK126 ($n = 6$ biological replicates for DZNep, $n = 3$ biological replicates for GSK126; $*P < 0.05$, $**P < 0.008$). **b**, Average percentage of anaphase structures with bridges ± s.e.m. in vehicle-, DZNep- or GSK126-treated cell lines ($n \geq 3$ biological replicates and indicated in Methods; $*P < 0.04$, $**P < 0.02$). **c**, Average percentage annexin V^{+}/$7AAD^{-}$ cells ± s.e.m. quantified by flow cytometry on cell lines cultured with indicated treatments for 3 days ($n = 4$ biological replicates; $*P < 0.03$, $**P < 0.001$ for etoposide versus dual-treated or DZNep versus dual treatment for PC9). **d**, 7AAD cell cycle flow cytometry was performed on cell lines with or without 5 μM etoposide or 1 μM DZNep for 4 days. The average percentage S phase ± s.e.m. of each culture is plotted ($n = 3$ biological replicates for HCC15, A549, PC9 and H23, $n = 4$ biological replicates for H157 and Sw1573; $*P < 0.05$ $**P < 0.009$).

S phase in response to DZNep or shEZH2 compared with treatment with etoposide alone, suggesting these lines undergo cell cycle arrest sparing them from apoptosis. In contrast, sensitized etoposide-treated lines had an average of 16.5% more cells in S phase in response to DZNep or shEZH2 (Fig. 3d and Extended Data Fig. 6a), suggesting that apoptosis of G2/M cells that could not repair anaphase bridges contributed to enrichment for S phase.

While the links between EZH2, BRG1 and TopoII explained increased sensitivity of *BRG1* mutants to TopoIIi, it was still unclear why *EGFR* mutants behaved similarly. In our panel of cell lines, 30% had a *BRG1* mutation and 23% had an *EGFR* mutation; however, none had both *BRG1* and *EGFR* mutations (Supplementary Table 4 and Extended Data Fig. 7a; Fisher's exact test $P = 0.005$). Similarly, in a panel of 412 sequenced primary adenocarcinomas[16,17], 65 (15.7%) had mutation in *EGFR*, while 33 (8%) had mutations in *BRG1*. With these allele frequencies, 1.25% of the tumour samples were expected to have both *BRG1* and *EGFR* mutations; however, only 1 (0.2%) was observed (Fisher's exact test, $P = 0.019$, EGFR R574L). The negative correlation of *BRG1* and *EGFR* mutations suggests that they may be functionally redundant—BRG1 loss may be permissive for high EGFR expression. Supporting this idea, a strong negative correlation was found between *EGFR* and *BRG1* expression in the Director's Challenge data set (Fig. 4a). When comparing gene expression of *EGFR* and *BRG1* mutant cell lines to WT lines, *EGFR* was the first of 21 shared upregulated genes, overexpressed 3.1-fold in *BRG1* mutant cell lines and 3.25-fold in *EGFR* mutant cell lines (Extended Data Fig. 7b, c and Supplementary Table 5).

To understand their genetic relationship better, EGFR and BRG1 expressions were manipulated (Fig. 4b and Extended Data Fig. 8a). BRG1 re-expression in the HCC15 lines converted the line from a sensitized to protected phenotype, and the HCC15 shEGFR line was no longer sensitized to etoposide by DZNep (Fig. 4c). Similarly, DZNep had no effect on etoposide IC_{50} when EGFR was overexpressed or BRG1 was knocked down in H460 cells (Fig. 4c). These changes in etoposide sensitivity were

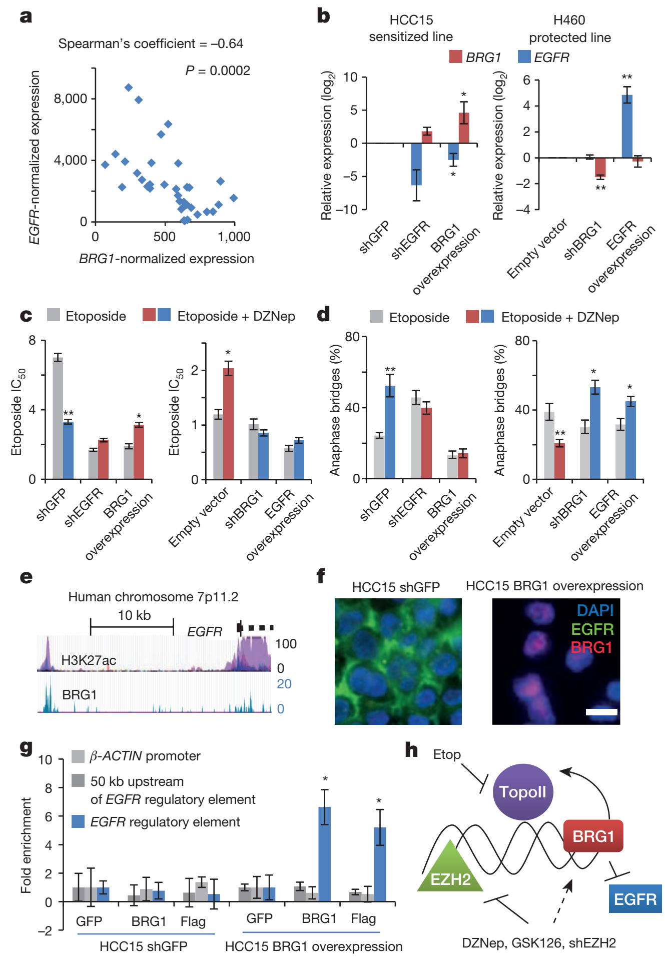


Figure 4 | BRG1 and *EGFR* are genetically antagonistic and control the sensitized phenotype. a, *EGFR* and *BRG1* expression in tumours from the Director's Challenge were plotted and correlation was assessed ($n = 36$). **b**, RT-qPCR for average expression of *BRG1* and *EGFR* ± s.e.m. in indicated transduced HCC15 and H460 cell lines ($*P < 0.04$, $**P < 0.0001$; $n = 5$ biological replicates for H460, $n = 4$ biological replicates for HCC15). **c**, Average etoposide IC_{50} ± s.e.m. in the indicated vehicle- and DZNep-treated cells ($n = 3$ biological replicates; $*P < 0.04$, $**P < 0.0001$). **d**, Average percentage of anaphase structures with bridges ± s.e.m. in indicated vehicle- or DZNep-treated cell lines ($n = 3$ biological replicates; $*P < 0.04$, $**P < 0.02$). **e**, Depiction of EGFR regulatory element including H3K27ac and BRG1 binding peaks and from ENCODE database. **f**, Immunofluorescence for EGFR and BRG1 on HCC15 cells transduced with short hairpin targeting green fluorescent protein (shGFP, control) and BRG1-overexpressing HCC15 cells; scale bar, 30 μm. **g**, Chromatin immunoprecipitation for GFP (control), BRG1 or the Flag-tag on the exogenously expressed BRG1 in the HCC15 shGFP (control) and BRG1-expressing cell lines. qPCR was performed with primers for indicated genomic sites and average enrichment over GFP ± s.e.m. was plotted ($n = 4$ biological replicates, one-way analysis of variance; $*P = 0.02$). **h**, In response to EZH2i: WT cells show increased BAF complex function and subsequent decrease in etoposide sensitivity; *BRG1* mutant cells have increased etoposide sensitivity; and EGFR mutant cells act like *BRG1* mutant cells owing to genetic antagonism of *EGFR* by BRG1.

consistent with changes in anaphase bridging, apoptosis and S phase accumulation (Fig. 4d and Extended Data Fig. 8b, c).

The ENCODE database[18] shows a chromatin immunoprecipitation peak for BRG1 upstream of *EGFR* (Fig. 4e). HCC15 cells re-expressing BRG1 had lower levels of EGFR by RT-qPCR and immunofluorescence (Fig. 4b, f). We hypothesized that BRG1-containing BAF complexes bind to the *EGFR* regulatory element to disrupt *EGFR* transcription, as observed in other systems[19]. We confirmed the exogenously expressed BRG1 was being incorporated into BAF complexes (Extended Data Fig. 9a), and then performed chromatin immunoprecipitation assays. BRG1, immunoprecipitated with an antibody recognizing the protein itself or the Flag-tag, was significantly associated with the *EGFR* regulatory element (Fig. 4g). This genetic antagonism explains why dividing cells in *EGFR*-driven cultures may not sufficiently upregulate BRG1, and respond to combined etoposide and EZH2i with increased anaphase bridging as *BRG1* mutant cells do (Fig. 4h).

Our results suggest that dual EZH2i plus TopoIIi represents a treatment option for *EGFR* mutant tumours, even those that invariably develop resistance to EGFR tyrosine-kinase inhibitors[12]. Furthermore, combination EZH2i plus TopoIIi offers the first specific therapeutic for *BRG1* mutant lung cancers. Importantly, the strong antagonism of etoposide and EZH2i warns against using this drug combination for EGFR and BRG1 WT tumours. Also notable is the fact that EZH2i either sensitized to etoposide (synergism) or protected from it (antagonism), but very rarely was additivity observed. We and others[20,21] have found that *BRG1* and *EGFR* mutations are significantly anti-correlated in NSCLC. Moreover, *BRG1* mutant NSCLCs had elevated *EGFR* levels, raising the possibility that *BRG1* mutant tumours represent a subset of patients with WT EGFR that respond to EGFR tyrosine-kinase inhibitor[20,21]. Unlike *Brg1*-deficient MEFs[14], *BRG1* mutant lung cancer cells did not exhibit TopoII dysfunction unless EZH2 was inhibited, suggesting a novel connection between EZH2 and TopoII function.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 November 2013; accepted 26 November 2014.
Published online 28 January 2015.

1. Jemal, A. *et al.* Global cancer statistics. *CA Cancer J. Clin.* **61,** 69–90 (2011).
2. Zornosa, C. *et al.* First-line systemic therapy practice patterns and concordance with NCCN guidelines for patients diagnosed with metastatic NSCLC treated at NCCN institutions. *J. Natl. Compr. Canc. Netw.* **10,** 847–856 (2012).
3. Wang, L. *et al.* Randomized phase II study of concurrent cisplatin/etoposide or paclitaxel/carboplatin and thoracic radiotherapy in patients with stage III non-small cell lung cancer. *Lung Cancer* **77,** 89–96 (2012).
4. Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome – biological and translational implications. *Nature Rev. Cancer* **11,** 726–734 (2011).
5. Simon, J. A. & Lange, C. A. Roles of the EZH2 histone methyltransferase in cancer epigenetics. *Mutat. Res.* **647,** 21–29 (2008).
6. Shedden, K. *et al.* Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nature Med.* **14,** 822–827 (2008).
7. Tan, J. *et al.* Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. *Genes Dev.* **21,** 1050–1063 (2007).
8. Choudhury, S. R. *et al.* (-)-Epigallocatechin-3-gallate and DZNep reduce polycomb protein level via a proteasome-dependent mechanism in skin cancer cells. *Carcinogenesis* **32,** 1525–1532 (2011).
9. McCabe, M. T. *et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492,** 108–112 (2012).
10. Chou, T. & Talalay, P. Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs of enzyme inhibitors. *Adv. Enzyme Regul.* **22,** 27–55 (1984).
11. Deweese, J. E. & Osheroff, N. The DNA cleavage reaction of topoisomerase II: wolf in sheep's clothing. *Nucleic Acids Res.* **37,** 738–748 (2009).
12. Ji, H. *et al.* The impact of human EGFR kinase domain mutations on lung tumorigenesis and *in vivo* sensitivity to EGFR-targeted therapies. *Cancer Cell* **9,** 485–495 (2006).
13. Jackson, E. L. *et al.* The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res.* **65,** 10280–10288 (2005).
14. Dykhuizen, E. C. *et al.* BAF complexes facilitate decatenation of DNA by topoisomerase II. *Nature* **497,** 624–627 (2013).
15. Wilson, B. G. *et al.* Epigenetic antagonism between Polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell* **18,** 316–328 (2010).
16. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150,** 1107–1120 (2012).
17. TCGA. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511,** 543–550 (2014).
18. ENCODE. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2013).
19. Hargreaves, D. C. & Crabtree, G. R. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* **21,** 396–420 (2011).
20. Oike, T. *et al.* A synthetic lethality-based strategy to treat cancers harboring a genetic deficiency in the chromatin remodeling factor BRG1. *Cancer Res.* **73,** 5508–5518 (2013).
21. Matsubara, D. *et al.* Lung cancer with loss of BRG1/BRM, shows epithelial mesenchymal transition phenotype and distinct histologic and genetic features. *Cancer Sci.* **104,** 266–273 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Kim laboratory, F. Luo, P. Louis, K. Harrington, X. Wang and J. Brainson for technical assistance and discussions, and J. Crabtree, D. Hargreaves, C. Kadoch, L. Zon, K. Cichowski, M. Enos, S. Orkin, A. Gutierrez and C. Roberts for discussions. This work was supported in part by the Ladies Auxiliary to the Veterans of Foreign Wars, PF-12-151-01-DMC from the American Cancer Society, and the Uniting Against Lung Cancer Young Investigator Award supported by Meryl Bralower (C.M.F.), Boston University Undergraduate Research Opportunities Program (P.T.D.), RO1 HL090136, U01 HL100402 RFA-HL-09-004, American Cancer Society Research Scholar Grant RSG-08-082-01-MGO, the V Foundation for Cancer Research, a Basil O'Conner March of Dimes Starter Award, the Harvard Stem Cell Institute, and the Lung Cancer Research Foundation (C.F.K.), the National Institutes of Health (NIH) grants CA122794, CA140594, CA163896, CA166480, CA154303 and CA120964 (K.K.W.), the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research (V.E.M.), and the NIH grant K08 CA163677 (P.S.H.).

Author Contributions C.M.F., C.X., K.K.W. and C.F.K. designed the study; C.M.F., C.X., P.T.J., J.M.B. and Y.J.L. performed the experiments; S.P.R. cloned the EZH2 complementary DNA (cDNA) vector; H.Z. performed EZH2 immunohistochemistry, V.E.M. provided DZNep, P.S.H. analysed primary tumour sequencing data; K.K.W. allowed autochthonous mouse models studies in his laboratory; C.M.F. and C.F.K. wrote the manuscript with comments from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.F.K. (carla.kim@childrens.harvard.edu).

METHODS

Cell lines. Cell lines used are listed in Supplementary Table 4. All cell lines were maintained in RPMI 1640 media with 10% fetal bovine serum, 4 mM L-glutamine and penicillin/streptomycin at 37 °C, 5% CO_2. Cell lines were obtained from the Meyerson laboratory at Dana-Farber Cancer Institute. No mycoplasma was detected in cultures by either routine mycoplasma PCR or perinuclear 4′,6-diamidino-2-phenylindole (DAPI) staining. Cell line genotypes (Supplementary Table 4) were obtained from published studies[22,23], COSMIC database[24] and CCLE database[25].

Vectors. The pLKO.1 EZH2 shRNA construct clones TRCN0000040076 and TRCN0000040073 were purchased from Sigma and the shGFP plasmid 12273 is available on Addgene[26]. Both shBRG1 and the matched empty vector were provided by the Smale laboratory[27] and are available on Addgene, the BRG1 overexpression plasmid 19148 from the Massagué laboratory[28] was purchased through Addgene, and the shEGFR and EGFR WT overexpression constructs were provided by the Jänne laboratory[29]. The EZH2 overexpression construct was derived by cloning human EZH2 cDNA into pLenti7.3/V5-DEST (Invitrogen). Lentivirus was packaged in HEK293T cells using established protocols[30], and retrovirus was packaged in PlatE cells again using established protocols[31]. Cell lines were infected with viral-containing supernatant containing 6 μg ml^{-1} polybrene (Sigma) for a period of 10–18 h. Infected cultures were selected with 1 μg ml^{-1} puromycin (all sh constructs and EGFR overexpression, SIMGA), 200 μg ml^{-1} hygromycin (BRG1 overexpression, Invitrogen), or by flow cytometry for GFP (EZH2 overexpression) 5 days after infection.

Small hairpin sequences: GFP: GCCC(GCAAGCTGACCCTGAAGTTCAT)TCAAGAG(ATGAACTTCAGGGTCAGCTTGC)TTTT; EZH2 coding region: CCGG(CGGAAATCTTAAACCAAGAAT)CTCGAG(ATTCTTGGTTTAAGATTTCCG)TTTTT; EZH2 3′ UTR: CCGG(TATTGCCTTCTCACCAGCTGC)CTCGAG(GCAGCTGGTGAGAAGGCAATA)TTTTT; EGFR: CCGG(GCTGAGAATGTGGAATACCTA)CTCGAG(TAGGTATTCCACATTCTCAGC)TTTTT; BRG1: TTTG(TGGATAAGCAGCACAAGATT)TCAAGAG(AATCTTCTGCTGCTTCTCCA)TTTTT.

Drugs. Etoposide and doxorubicin (Sigma) were diluted to a stock of 100 mM in dimethyl sulphoxide (DMSO) for all cell culture experiments. DZNep was a gift from V.E.M. and was diluted in DMSO to a stock of 10 mM. GSK126 was purchased from Xcess Bio as a 10 mM stock in DMSO. All stocks were diluted in DMSO to 1,000× concentration before addition into media at 2× concentration and final dilution onto plated cells 1:1.

Cytotox assays. Cell lines were dissociated, counted and plated at 5,000 cells per well in flat-bottomed opaque tissue-culture-treated 96-well plates (CytoOne). Edge wells were filled with PBS. The following day, 2× drug diluted in media was added to each well such that the well then contained 100 μl media with 1× drug concentration at the following doses: etoposide, 0, 0.1, 1, 3, 5, 7, 10, 50, 100, 500 μM; or doxorubicin, 0, 0.01, 0.1, 0.5, 1, 3, 5, 7, 10, 50 μM; with or without additional 1 μM DZNep, 10 μM GSK126 or a continuation of 2 μM GSK126 from a 9-day 2 μM GSK126 pretreated culture. After 4 days, CellTiter-Glo (Promega) was added and luminescence was read on a BioTec plate reader to determine relative cell number in each well. Data were averaged for triplicate or quadruplicate technical replicates and normalized to the untreated wells, and whole runs (vehicle and EZH2i) were excluded if vehicle-treated wells did not reach the threshold luminescence of 3,000. Results from independent biological replicate experiments were input into GraphPad Prism software to extrapolate the IC_{50} and s.e.m. of IC_{50} for a given cell line using the nonlinear regression analysis of log(inhibitor) versus normalized response with a variable slope. For 1 μM DZNep for 4 days with etoposide, $n = 3$ biological replicates for H1975, H2030, HCC4006, A549, HCC2450, Calu1, H1650, H522, H2126, H1299, HCC15, H322, H2009, HCC95, H520, H460, Calu3, H2122, H23 and H3255; $n = 4$ biological replicates for PC9, H157, HCC827, Sw1573, Calu6 and H441. For both 2 μM GSK126 pretreated for 9 days and continued for 4 days with etoposide, and 10 μM GSK126 treated for 4 days with etoposide, $n = 3$ biological replicates. The log(IC_{50}) values were compared using GraphPad Prism software, and P values reported are the sum-of-squares F statistics. For graphs, log(IC_{50}) of vehicle control cells was subtracted from log(IC_{50}) of EZH2i-treated cells and multiplied by ten to be depicted as log fold change × 10. Errors were estimated by calculating possible upper and lower bounds of log fold changes based on GraphPad reported s.e.m. for each log(IC_{50}) calculated.

For Chou–Talalay combination index assays[10], doses for etoposide, DZNep or GSK126 were 0, 0.1, 0.5, 1, 3, 5, 7, 10, 30, 50 μM and 1:1 combinations of etoposide and DZNep or etoposide and GSK126. Survival percentages for three independent biological replicate experiments were averaged and input into CompuSyn software to extrapolate combination index values. Any drug dose with mean survival over 100% was excluded because the CompuSyn software did not allow for values over 1. In addition, visual inspection led us to remove the highest doses of GSK126 from the analysis for every cell line, as suggested in refs 32 and 33, leading to much better matching of the data points to the median effect plot. Notably, removal of these data points both increased synergy seen in the sensitized lines and increased antagonism seen in H460 and Sw1573 lines.

Flow cytometry. For 7AAD-cell cycle analysis, cell lines were plated at 1.5×10^6 cells per 10 cm plate and treated with drug for 4 days. Cells were then dissociated, fixed with 100% ice cold ethanol for at least 2 h, incubated for 30 min with 1 mg ml^{-1} DNase-free RNase A (Thermo) and resuspended in 20 μg ml^{-1} 7-aminoactinomycin D (7AAD; Invitrogen). Thirty thousand events were collected on the BD Fortessa, analysed with ModFit LT software and the results were averaged for three or four biological replicates (n indicated in legend).

For Annexin V/7AAD apoptosis analysis, cell lines were plated at 5×10^4 cells per well of a six-well plate and treated with drug for 3 days. Supernatant was retained and added to trypsinized suspensions of adherent cells. Cells were stained with Annexin V-FITC (BD Biosciences) according to the manufacturer's instructions, and resuspended with 1 μg ml^{-1} 7AAD before analysis on BD Fortessa. For sensitized lines 2 μM GSK126 pretreated for 9 days and continued for 3 days with etoposide was used, while for protected line 10 μM GSK126 treated for 3 days with etoposide was used. Data were analysed with FlowJo (Treestar) software and the percentages of Annexin V^+/7AAD$^-$ cells were averaged for four biological replicate experiments.

Quantitative RT–PCR. RNA from treated cell lines was extracted using Absolutely RNA kits (Agilent) and cDNA was made using the SuperScript III kit (Invitrogen). Relative gene expression was assayed with Sybr green on the StepOnePlus real-time PCR system (Applied Biosystems). Relative expression was calculated by using threshold cycle (C_t) values: gene of interest($C_{t,\ reference} - C_{t,\ experimental}$) − CYPA($C_{t,\ reference} - C_{t,\ experimental}$) and graphs potted on the $\log_2$ scale or converted to linear scale. Statistics were performed on $\log_2$ data. For all experiments, the reference sample was a matched vehicle-treated or control transduced cell line.

Primer sequences: CYPA: F TCATCTGCACTGCCAAGACTG R CATGCCTTCTTTCACTTTGCC; EZH2: F AGGAGTTTGCTGCTGCTCTC R CCGAGAATTTGCTTCAGAGG; BRG1: F AGCGATGACGTCTCTGAGGT R GTACAGGGACACCAGCCACT; EGFR: F TAACAAGCTCACGCAGTTGG R GTTGAGGGCAATGAGGACAT.

Xenograft experiments. For DZNep experiments, H157 or H23 cells were dissociated into single cells, counted and resuspended at 1×10^6 cells per 250 μl of 1:1 media/matrigel (BD). Eight- to 16-week-old female Foxn1^{nu}/Foxn1^{nu} (Nude) mice (Harlan) were injected subcutaneously with 1×10^6 cells in two to four spots on flanks. Etoposide and DZNep were administered from day 12 to day 17 after injections; etoposide: 20 mg/kg/d intraperitoneally in corn oil once per day for 5 consecutive days; DZNep 2 mg/kg/d intraperitoneally in corn oil twice per week for 1 week, or 1 mg/kg/d intraperitoneally in corn oil twice per week for 2 weeks. Tumour growth was measured every other day by calliper in a non-blinded fashion. For GSK126 experiments, PC9 cells were dissociated into single cells, counted and resuspended at 1×10^6 cells per 250 μl of 1:1 media/matrigel (BD). Eight- to 16-week-old female Foxn1^{nu}/Foxn1^{nu} mice (Harlan) were injected subcutaneously with 1×10^6 cells in one spot on left flank. Tumours were allowed to grow for 23 days to a mean size of 70 mm^3. Mice were then randomized into groups that received etoposide, GSK126, both etoposide and GSK126, gefitinib or vehicle: etoposide 10 mg/kg/d intraperitoneally in corn oil three times a week, GSK126 300 mg/kg/d intraperitoneally in 1:1 v/v GSK126/Captisol mixture resuspended in sterile water with acetic acid to pH 4.8, gefitinib (LC Laboratories) 150mg/kg/d in 1% Tween-80 (Sigma). All mouse experiments were approved by the BCH Animal Care and Use Committee and by the Dana-Farber Cancer Institute Institutional Animal Care and Use Committee, both accredited by the Association for Assessment and Accreditation of Laboratory Animal Care, and were performed in accordance with relevant institutional and national guidelines and regulations.

Generation of the EZH2 co-expression gene signature. We used Oncomine[34] to query the top 20 genes co-expressed with EZH2 in all data sets containing human non-small-cell lung cancer samples and co-expression data[6,35–41]. We chose 20 probes for examination from each study, to yield a list between 100 and 200 genes, which allowed for robust hierarchal clustering of samples similar to that in previous studies. Of the 180 probes, 64 were redundant, leading to a list of 116 genes highly co-expressed with EZH2 (Supplementary Table 1). Because these data sets were from various microarray platforms, the gene list was then used to generate a probe list for the 116 genes corresponding to probes on the U133A Affymetrix array using the batch query function on the NetAffx website (http://www.affymetrix.com/analysis/index.affx). Gene ontology analysis was performed on the *EZH2* co-expression signature with dChip software (http://www.hsph.harvard.edu/cli/complab/dchip/).

Microarray analysis. All array data are publically available on Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) and correspond to array files available from GSE4824 (ref. 42) for all lines except A549, H522 and PC9, GSE5457 for A549 (two replicates), GSE5720 (ref. 43) for H522 and an additional A549, GSE7670 (ref. 44) for an additional H1299, GSE10089 (ref. 45) for PC9 and H1650, and

GSE31625 (ref. 46) for an additional PC9. Arrays were chosen based on availability in September 2012. Arrays were analysed using R/Bioconductor (http://www.bioconductor.org/). Raw CEL files from U133A Affymetrix arrays were processed using the robust multiarray average algorithm[47]. To identify genes correlating with the phenotypic groups, we used limma[48] to fit a statistical linear model to the data and then tested for differential gene expression in the three groups using the vennSelect package—WT: H460, H441, H2122, H2009, Calu6, HCC95; EGFR mutant: H1650, HCC827, HCC4006, H1975, H3255, PC9; BRG1 mutant: A549, H1299, H157, H2126, H522, HCC15. Results were adjusted for multiple testing using the Benjamini and Hochberg method[49], and significance was determined using a false-discovery-rate cutoff of less than 5%. For correlation between *EGFR* and *BRG1* expression in the Director's Challenge data set, the arrays were processed using RMA and limma as described above. Tumours with a robust multiarray average normalized expression of more than 2,000 for *EGFR* or 575 for *BRG1* were plotted and correlation was assessed—because these data were nonparametric, Spearman's correlation coefficient was used. Using this same method of selecting for highest-expressing tumours, we could visualize positive correlations between *EZH2* and other *EZH2* co-expression gene signature members (data not shown).

Kaplan–Meier analysis. Raw gene expression data from the Director's Challenge human lung adenocarcinoma samples[6] were obtained (https://caintegrator.nci.nih.gov/caintegrator/). Probe intensities from the Affymetrix U133A platform used in these studies were normalized and modelled using dChip software[50] (http://www.hsph.harvard.edu/cli/complab/dchip/). Kaplan–Meier survival analyses were implemented after the samples were hierarchically clustered using centroid linkage, rank correlation and gene peaking time into two risk groups using the EZH2 co-expression gene signature. Survival differences between the two risk groups were assessed using the Mantel–Haenszel log-rank test. The large area between the two risk groups and its associated small *P* value from the Mantel–Haenszel log-rank test implicate a robust survival classification model.

Statistical analysis. No statistical methods were used to predetermine sample size. Except where indicated, a two-tailed Student's *t*-test with equal variance was used to compare measurements between two conditions with at least three biological replicates per condition. Normal distribution was checked with the Kolmogorov–Smirnov test; data that failed the test at $\alpha = 0.05$ were considered normally distributed. If the data were non-parametric, then a two-tailed Mann–Whitney *U*-test was used instead of a *t*-test to assess the *P* value. Unless noted otherwise, pooled data are represented by the mean and standard error. *P* values are indicated in figure legends, and *P* values less than 0.05 were considered significant.

Western blot. Whole-cell extracts were made in RIPA buffer (0.5% Deoxycholate, 1% IGEPAL-CA630, 0.1% sodium dodecyl sulphate, 150 mM NaCl, 50 mM Tris-8.1), lysates were cleared by centrifugation, and protein concentrations were quantified with the Pierce BCA Protein Assay Kit (Thermo). For western blotting, 25 μg of protein extract per sample was denatured with heat and reducing agents, separated on a 4–12% acrylamide gel (BioRad) and transferred to nitrocellulose (GE Healthcare). Antibodies used for western blotting were as follows: EZH2 (clone D2C9; Cell Signaling; 1:200), Histone H3 (polyclonal; AbCAM ab1791; 1:2,000) and H3K27me3 (polyclonal; Millipore 07-449; 1:1,000), all incubated overnight at 4 °C. All antibodies have detailed species validation available online from vendors. Secondary antibody, anti-rabbit-HRP (Santa Cruz sc-2313; 1:10,000), was incubated for 1 h at room temperature. After washing, chemiluminescence was visualized with Western Lightning Plus-ECL (PerkinElmer) and exposure onto KODAK BioMax XAR film.

Immunoprecipitation. Cultured cells (10×10^6 per line) were collected by trypsinization and pelleting, followed by PBS wash and pelleting. On ice, cell pellets were resuspended in 750 μl hypotonic buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$) and incubated for 15 min. NP-40 (10%, 37.5 μl) was added to each tube followed by vortexing for 10 s and centrifugation for 10 min at 650*g* at 4 °C. Nuclear pellets were then resuspended in RIPA buffer for 30 min on ice with vortexing, lysates were cleared by centrifugation and protein concentrations were quantified with the Pierce BCA Protein Assay Kit (Thermo). Antibodies directed against Flag (M2; Sigma; 1:50), BAF155 (R-18; Santa Cruz; 1:50) or ARID1a (PSG3; Santa Cruz; 1:10) were incubated with 300 μg of each nuclear extract with 1:1 Protein A and Protein G agarose beads (GE Healthcare) rotating at 4 °C overnight. All antibodies had detailed species validation available online from vendors. Beads were washed three times in 1 ml RIPA buffer with the second wash being rotated at 4 °C for 30 min, then resuspended in approximately 35 μl 1× reducing buffer and boiled for 5 min before loading 10 μl per lane. Proteins were separated on a 4–12% acrylamide gel (BioRad) and transferred to nitrocellulose (GE Healthcare). Antibodies for western blotting were as follows: Flag (M2; Sigma; 1:1,000), BAF155 (R-18; Santa Cruz; 1:1,000) and ARID1a (polyclonal; Bethyl Laboratories A310-040A; 1:1,000). Secondary antibody, anti-rabbit-HRP (Santa Cruz sc-2313; 1:10,000), anti-mouse-HRP (Santa Cruz sc-2314; 1:10,000) or anti-goat-HRP (Santa Cruz sc-2020; 1:10,000) were incubated for 1 h at room temperature. After washing, chemiluminescence was visualized with Western Lightning Plus-ECL (PerkinElmer) and exposure onto KODAK BioMax XAR film.

Anaphase bridge analysis. To quantify anaphase bridges, cells were grown on four-well cultures slides (Lab Tek II). Adherent cells were fixed with 4% paraformaldehyde for 20 min, washed and stained with Vectashield with DAPI (Vector Labs). Images were taken of each anaphase structure, and the number of anaphases with bridges over the total number of anaphases (between 11 and 34 total anaphases per well of a four-well chamber slide) was recorded for each of three or more independent biological replicate experiments in a blinded fashion. Exact biological replicates *n* for Fig. 3b were as follows: PC9 vehicle = 7, DZNep = 4, GSK126 = 4; A549 vehicle = 3, DZNep = 3, GSK126 = 3; H157 vehicle = 7, DZNep = 5, GSK126 = 3; H23 vehicle = 6, DZNep = 4, GSK126 = 3; Sw1573 vehicle = 4, DZNep = 4, GSK126 = 3; and for Extended Data Fig. 5b were H441 vehicle = 3 DZNep = 3; H2009 vehicle = 3, DZNep = 3; H522 vehicle = 3, DZNep = 3; and H1650 vehicle = 4, DZNep = 4. Imaging was performed with a Nikon 90i camera with ×100 objective and oil emersion and NIS-Elements software, and processed with NIS-Elements and Adobe Photoshop.

Immunofluorescence. Cells were fixed in 4% paraformaldehyde and permeabilized with 10% Normal Donkey Serum (Jackson Immunoresearch), 0.25% Triton-X (Sigma), both in PBS. Primary antibodies, Brg1 (clone G-7, Santa Cruz) and EGFR (polyclonal; Cell Signaling 2232) were incubated overnight at 1:100 dilution in PBS, 10% Normal Donkey Serum. All antibodies had detailed species validation available online from vendors. Slides were washed three times and secondary antibodies, anti-mouse-AlexaFluor594 and anti-rabbit-AlexaFluor488 (Invitrogen) were incubated at 1:500 for 1 h. After washing, cover slips were mounted with Vectashield with DAPI (Vector Labs). Imaging was performed with a Nikon 90i camera and NIS-Elements software, and processed with NIS-Elements and Adobe Photoshop. All treatment groups were imaged with the same exposure time and equivalent processing. Images were chosen to highlight the difference between BRG1^{high} interphase cells and EGFR^{high} dividing cells in EZH2i-treated PC9 cultures.

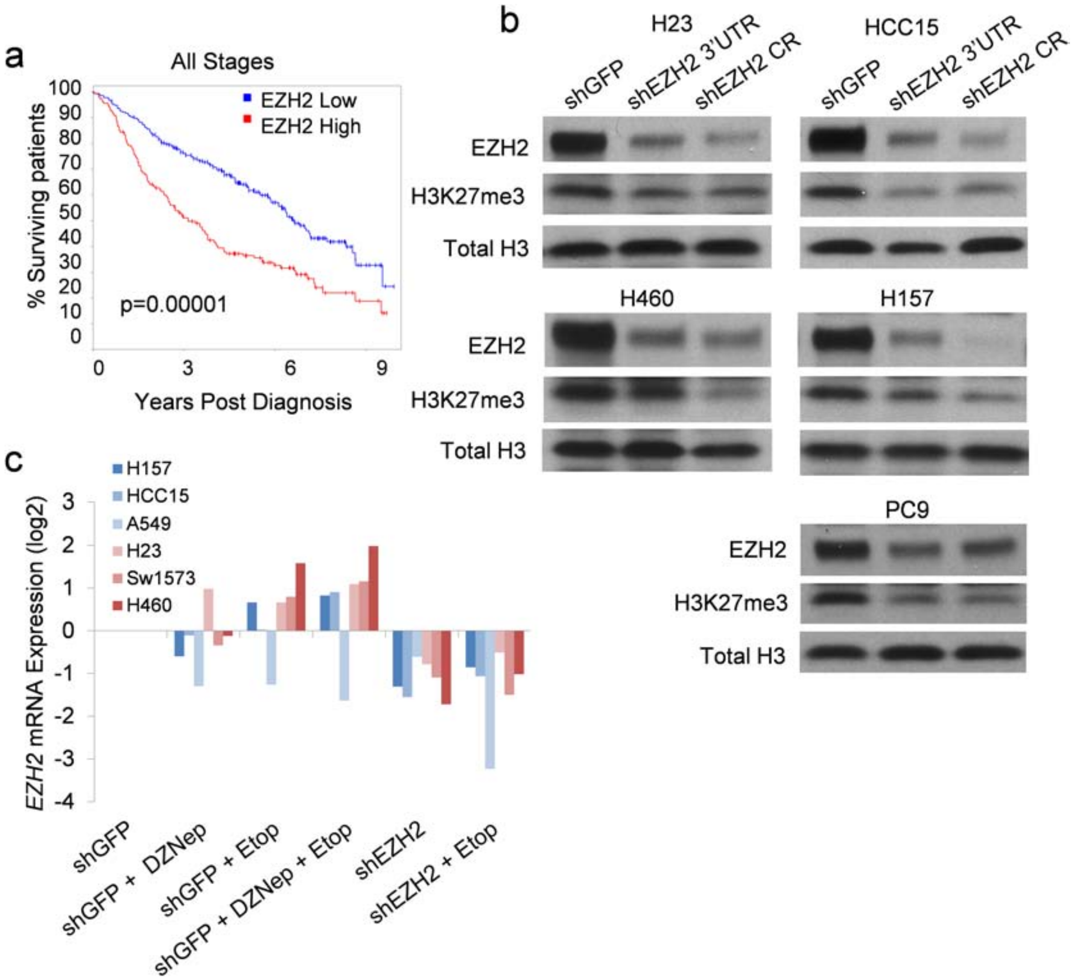
Treatment and magnetic resonance imaging of endogenous mouse models. Doxycycline-inducible *EGFR*$^{T790M;L858R}$ transgenic mice[12] and *Lox-Stop-Lox-Kras*$^{G12D/+}$*;p53*$^{fl/fl}$ (*Kras*$^{G12D/+}$*;p53*$^{\Delta/\Delta}$)[13,51] mice were maintained on a mixed background, housed in a pathogen-free environment at the Harvard School of Public Health and handled in strict accordance with Good Animal Practice as defined by the Office of Laboratory Animal Welfare. All animal work was done with Dana-Farber Cancer Institute Institutional Animal Care and Use Committee approval. Cohorts of male and female *EGFR*$^{T790M;L858R}$;CCSP-rtTA were put on a doxycycline diet at 6 weeks of age to induce the expression of mutant EGFR, while male and female *Kras*$^{G12D/+}$*;p53*$^{\Delta/\Delta}$ mice received intranasal adeno-Cre between 6 and 8 weeks of age. Mice were evaluated by magnetic resonance imaging 12–16 weeks after doxycycline diet or adeno-Cre infection to document and quantify the lung cancer burden before being randomized to various treatment study cohorts. Treated mice in all cohorts had similar initial tumour burden. Tumour-bearing mice were randomized into cohorts treated either with vehicle (corn oil), etoposide 10 mg/kg intraperitoneally three times a week for 4 weeks, DZNep 4 mg/kg intraperitoneally twice a week for 4 weeks or both etoposide and DZNep. The mice were imaged by magnetic resonance imaging every 2 weeks to determine the reduction in tumour volume during the respective treatments as described previously in a non-blinded fashion[12]. The tumour burden volume and quantification were reconstructed on three-dimensional slicer software (http://www.slicer.org). Immunohistochemistry was performed as described with anti-EZH2 (clone D2C9; Cell Signaling) or anti-pEGFR (Y1068; clone D7A5; Cell Signaling) and developed using Vectastain Elite ABC kit (Vector Labs). Imaging was performed with a Nikon 90i camera and NIS-Elements software, and processed with NIS-Elements and Adobe Photoshop.

Chromatin immunoprecipitation. Five million cells were fixed in 1% formaldehyde for 10 min before addition of glycine to a concentration of 1 mM. Cells were pelleted, washed and resuspended in chromatin immunoprecipitation sonication buffer (1% Triton X-100, 0.1% deoxycholate, 50 mM Tris 8.1, 150 mM NaCl, 5 mM EDTA) containing protease and phosphastase inhibitors (Roche). Samples were sonicated for a total of 3 min in 30 s cycles with 1 min breaks. Sonicated samples were centrifuged for 15 min to clear the lysates, and resulting whole-cell extracts were used for pull-downs. Antibodies directed against GFP (Ab-1; Neomarkers), BRG1 (G-7; Santa Cruz) and Flag (M2; Sigma) were incubated with equal proportions of whole-cell extracts at 1:30 dilution overnight, rotating at 4 °C. Protein A and Protein G agarose beads (1:1; GE Healthcare) were added and incubated for 2 h at 4 °C. Beads were then pelleted and washed with high salt wash buffer (1% Triton X-100, 0.1% deoxycholate, 50 mM Tris-8.1, 500 mM NaCl, 5 mM EDTA), followed by LiCl immune complex buffer (250 mM LiCl, 0.5% IGEPAL-CA630, 0.5% deoxycholate, 10 mM Tris-8.1, 1 mM EDTA) and TE (10 mM Tris-8.1, 1 mM EDTA) before suspension in elution buffer (1% SDS, 0.1 M $NaHCO_2$, 0.01 mg ml^{-1} salmon sperm DNA (GE Healthcare)). Crosslinks were reversed at 65 °C overnight, beads were pelleted, and resulting supernatant was incubated with 0.4 mg ml^{-1} Proteinase K (Sigma) for

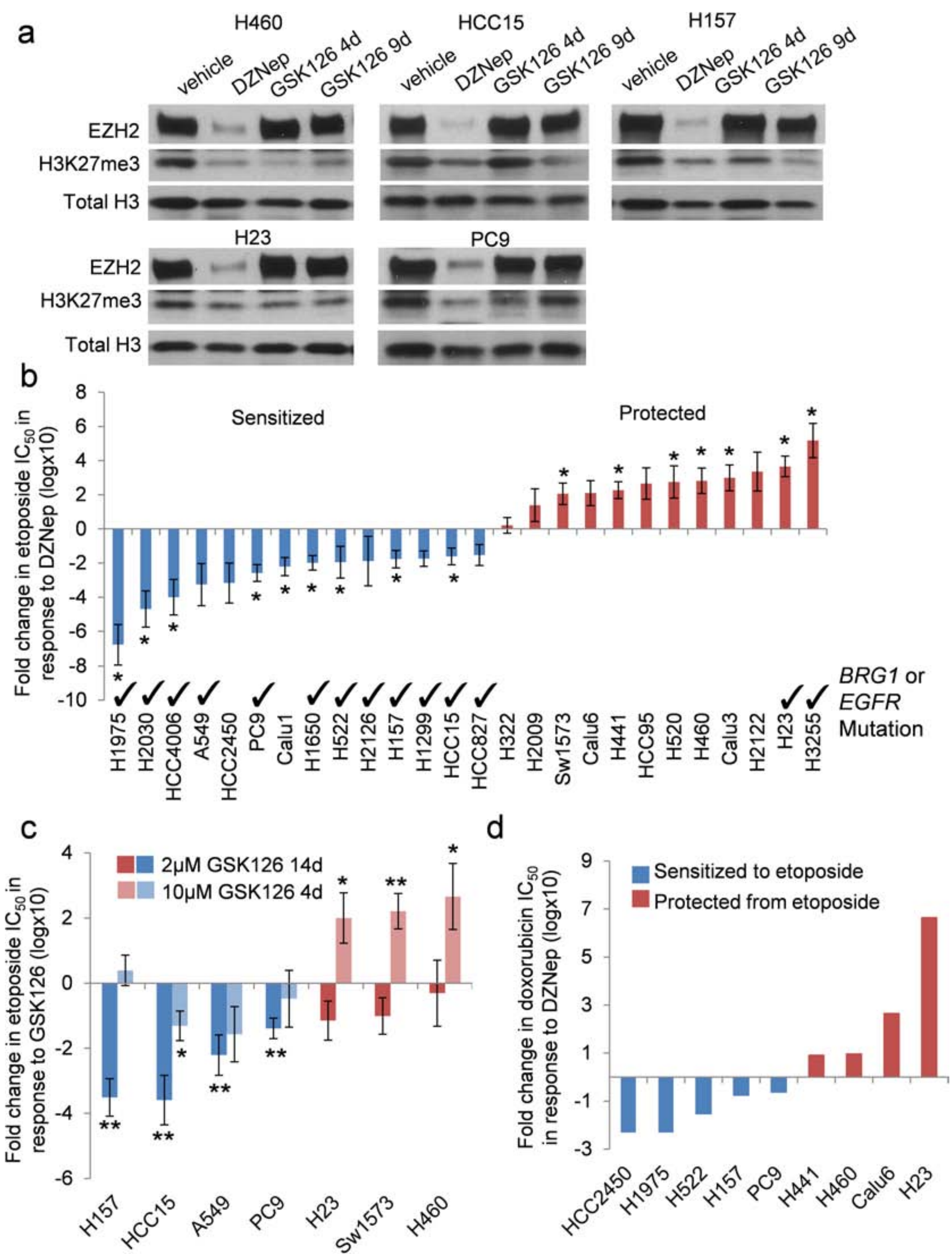
2 h at 37 °C. DNA from each sample was purified using Qiagen PCR purification columns following the manufacturer's instructions. Samples were resuspended in 100 μl 10 mM Tris-8.1 and 2 μl were used for each Sybr green PCR reaction (Applied Biosystems). Enrichment was calculated by average ($C_{t,\ reference} - C_{t,\ experimental}$) and converted to the linear scale for each genomic region of interest. Reference samples were the GFP; both BRG1 and Flag samples were experimental. Statistical analyses (one-way analysis of variance) were performed on $\log_2$-transformed data for three independent biological replicates.

Primer sequences: β-actin: F TCGAGCCATAAAAGGCAACT R TCTCCCTC CTCCTCTTCCTC; EGFR regulatory element: F CCTTGTAGATTGGGGCTGA G R AGTTTGGGGGTGGAAGAAAG; 50kb upstream of regulatory element: F GGCTGAGACAGAGGGAACAC R CCATCTCAGCCTCCCAAGTA.

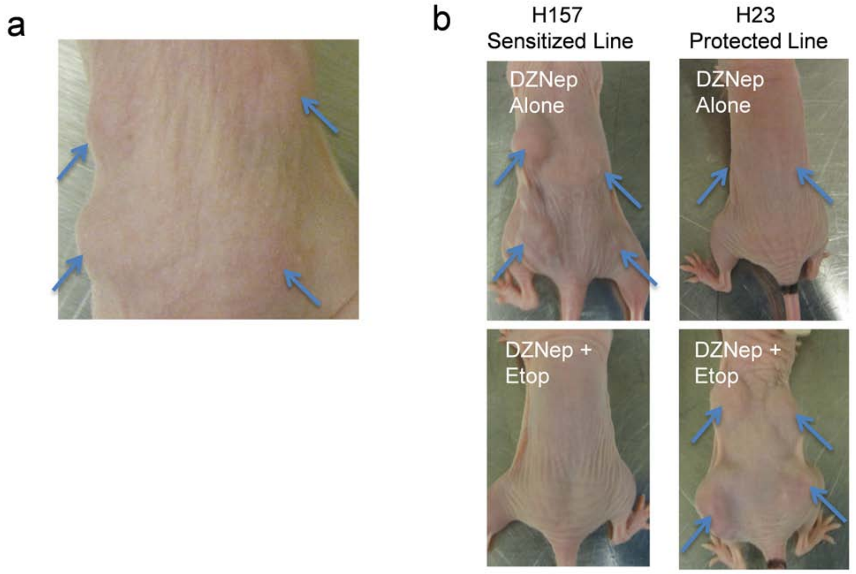
22. Medina, P. P. *et al.* Frequent BRG1/SMARCA4-inactivating mutations in human lung cancer cell lines. *Hum. Mutat.* **29,** 617–622 (2008).
23. Yamamoto, H. *et al.* PIK3CA mutations and copy number gains in human lung cancers. *Cancer Res.* **68,** 6913–6921 (2008).
24. Bamford, S. *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **91,** 355–358 (2004).
25. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–607 (2012).
26. Orimo, A. *et al.* Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 Secretion. *Cell* **121,** 335–348 (2005).
27. Ramirez-Carrozzi, V. R. *et al.* Selective and antagonistic functions of SWI/SNF and Mi-2 nucleosome remodeling complexes during an inflammatory response. *Genes Dev.* **20,** 282–296 (2006).
28. Xi, Q., He, W., Zhang, X. H.-F., Le, H.-V. & Massagué, J. Genome-wide impact of the BRG1 SWI/SNF chromatin remodeler on the transforming growth factor β transcriptional program. *J. Biol. Chem.* **283,** 1146–1155 (2008).
29. Engelman, J. A. *et al.* Allelic dilution obscures detection of a biologically significant resistance mutation in EGFR-amplified lung cancer. *J. Clin. Invest.* **116,** 2695–2706 (2006).
30. Fillmore, C. M. *et al.* Estrogen expands breast cancer stem-like cells through paracrine FGF/Tbx3 signaling. *Proc. Natl Acad. Sci. USA* **107,** 21737–21742 (2010).
31. Zacharek, S. J. *et al.* Lung stem cell self-renewal relies on bmi1-dependent control of expression at imprinted loci. *Cell Stem Cell* **9,** 272–281 (2010).
32. Chou, T.-C. Drug combination studies and their synergy quantification using the Chou-Talalay method. *Cancer Res.* **70,** 440–446 (2010).
33. Chou, T.-C. Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. *Pharmacol. Rev.* **58,** 621–681 (2006).
34. Rhodes, D. R. *et al.* Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9,** 166–180 (2007).
35. Beer, D. G. *et al.* Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Med.* **8,** 816–824 (2002).
36. Garber, M. E. *et al.* Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl Acad. Sci. USA* **98,** 13784–13789 (2001).
37. Gordon, G. J. *et al.* Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. *Cancer Res.* **62,** 4963–4967 (2002).
38. Landi, M. T. *et al.* Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS ONE* **3,** e1651 (2008).
39. Rohrbeck, A. *et al.* Gene expression profiling for molecular distinction and characterization of laser captured primary lung cancers. *J. Transl. Med.* **6,** 69 (2008).
40. Su, A. I. *et al.* Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Res.* **61,** 7388–7393 (2001).
41. Yu, K. *et al.* A precisely regulated gene expression cassette potently modulates metastasis and survival in multiple solid cancers. *PLoS Genet.* **4,** e1000129 (2008).
42. Lockwood, W. W. *et al.* DNA amplification is a ubiquitous mechanism of oncogene activation in lung and other cancers. *Oncogene* **27,** 4615–4624 (2008).
43. Shankavaram, U. T. *et al.* Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. *Mol. Cancer Ther.* **6,** 820–832 (2007).
44. Su, L.-J. *et al.* Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme. *BMC Genomics* **8,** 140 (2007).
45. Miyanaga, A. *et al.* Antitumor activity of histone deacetylase inhibitors in non-small cell lung cancer cells: development of a molecular predictive model. *Mol. Cancer Ther.* **7,** 1923–1930 (2008).
46. Balko, J. *et al.* Gene expression patterns that predict sensitivity to epidermal growth factor receptor tyrosine kinase inhibitors in lung cancer cell lines and human lung tumors. *BMC Genomics* **7,** 289 (2006).
47. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4,** 249–264 (2003).
48. Smyth, G. K., Michaud, J., I. & Scott, H. S. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21,** 2067–2075 (2005).
49. Hochberg, Y. & Benjamini, Y. More powerful procedures for multiple hypothesis testing. *Stat. Med.* **9,** 811–818 (1990).
50. Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proc. Natl Acad. Sci. USA* **98,** 31–36 (2001).
51. Curtis, S. J. *et al.* Primary tumor genotype is an important determinant in identification of lung cancer propagating cells. *Cell Stem Cell* **7,** 127–133 (2010).



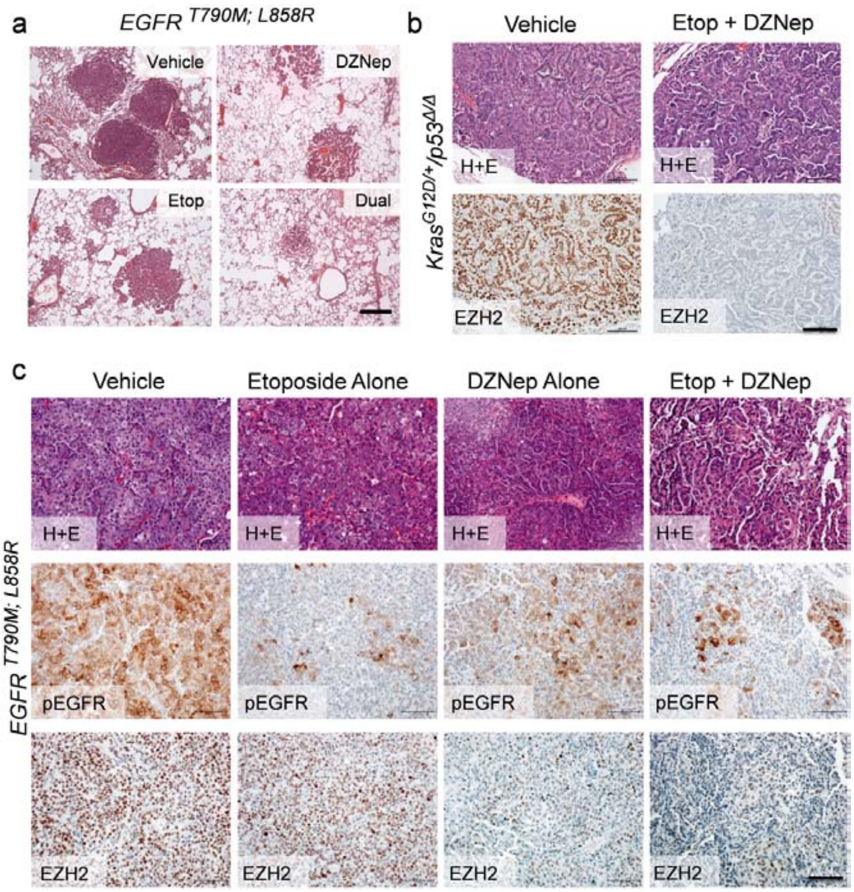
Extended Data Figure 1 | Verification of EZH2 as a potential target for NSCLC. **a**, Survival of patients with lung adenocarcinoma in the Director's Challenge data set. Samples were hierarchically clustered using the primary-tumour-generated *EZH2* co-expression signature (Supplementary Table 1) into two risk groups. The Kaplan–Meier curve for the whole data set is shown ($n = 416$, $P < 0.00001$). **b**, Western blot was performed on whole-cell extracts from indicated lines for EZH2 and its catalytic mark H3K27me3; total histone H3 is shown as loading control. CR indicates a coding region targeting hairpin. **c**, RT-qPCR for average expression of *EZH2* in the indicated cell lines after plating at equal density and treating for 4 days with indicated treatments. Each cell line is normalized to its shGFP control ($n = 2$ biological replicates).



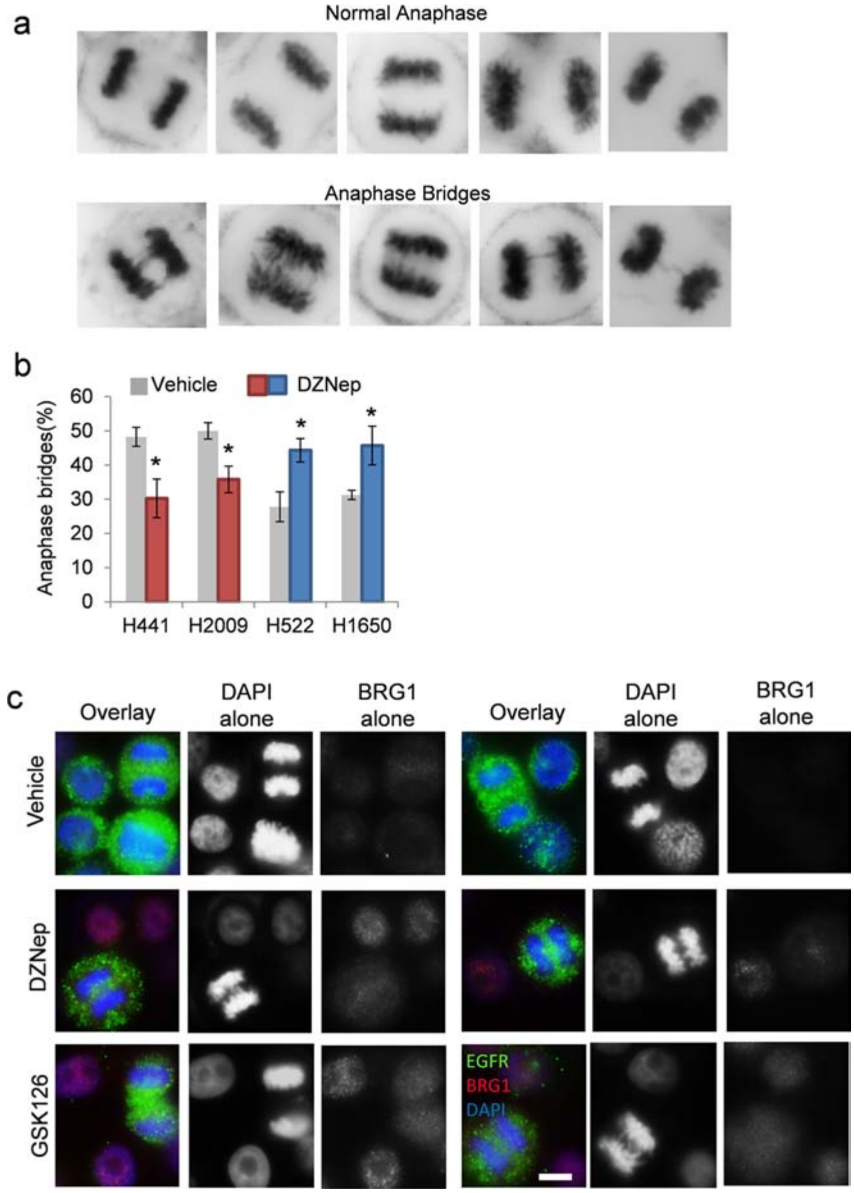
Extended Data Figure 2 | Pharmacological inhibition of EZH2 changes response of cells to TopoII inhibitors. **a**, Western blot for EZH2 and H3K27me3 was performed on whole-cell extracts after administration of 1 μM DZNep for 4 days, 10 μM GSK126 for 4 days, 2 μM GSK126 for 9 days, or vehicle. Total histone H3 is shown as a loading control. **b**, The fold change in etoposide IC_{50} ± s.e.m. in response to DZNep is plotted ($n = 3$ biological replicates for H1975, H2030, HCC4006, A549, HCC2450, Calu1, H1650, H522, H2126, H1299, HCC15, H322, H2009, HCC95, H520, H460, Calu3, H2122, H23 and H3255; $n = 4$ biological replicates for PC9, H157, HCC827, Sw1573, Calu6 and H441; $*P < 0.02$). Cell lines with mutations in *BRG1* or *EGFR* are indicated. Note that the H23 cell line has a very late coding region mutation in BRG1 (K1533N) and is predicted to produce functional protein[22], consistent with its protected phenotype in our assays. **c**, Fold change ± s.e.m. between vehicle-treated and indicated EZH2i-treated lines for etoposide IC_{50} is plotted ($n = 3$ biological replicates; $*P < 0.03$, $**P < 0.01$). **d**, Fold change in doxorubicin IC_{50} in response to DZNep ($n = 2$ biological replicates).



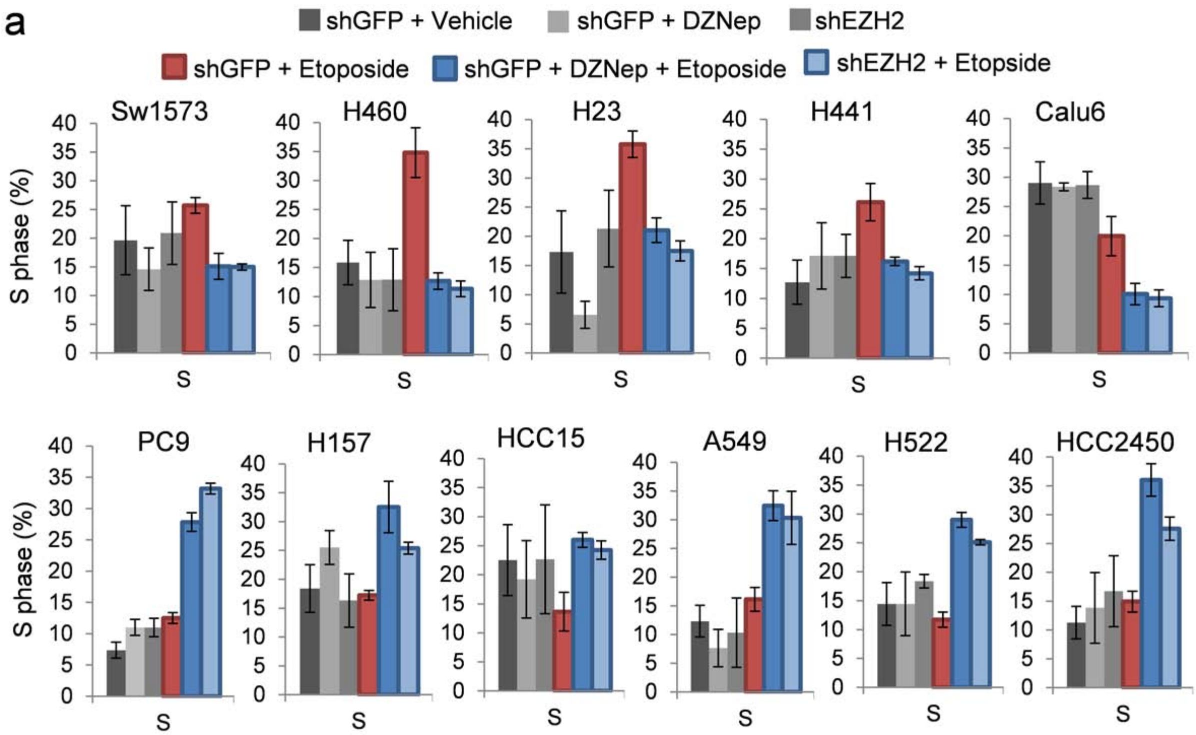
Extended Data Figure 3 | Xenograft experiments confirm sensitized and protected phenotypes. **a**, Representative image of mouse injected at four sites (arrows) with H23 tumour cells 12 days after cell injection. **b**, Representative images of mice injected at four sites with either H23 or H157 cells, and treated with indicated drugs, 35 days after cell injection. Palpable tumours that remain are indicated with arrows.



Extended Data Figure 4 | Autochthonous mouse models confirm genotype specificity of dual EZH2i and TopoIIi. **a**, Representative images of haematoxylin and eosin stained lung from *EGFR*$^{T790M;L858R}$ mice treated with indicated therapies for 4 weeks. Areas with tumours of similar sizes were chosen for comparison; scale bar, 200 μm. **b**, Histology from *Kras*$^{G12D/+}$/*p53*$^{\Delta/\Delta}$ mouse lung tumours after 1 week of indicated treatments; top image is haematoxylin and eosin, bottom image is EZH2 immunohistochemistry; scale bar, 100 μm. **c**, Histology from *EGFR*$^{T790M;L858R}$ mouse lung tumours after 4 weeks of indicated treatments; top image is haematoxylin and eosin, centre image is phospho-EGFR immunohistochemistry and bottom image is EZH2 immunohistochemistry; scale bar, 100 μm.



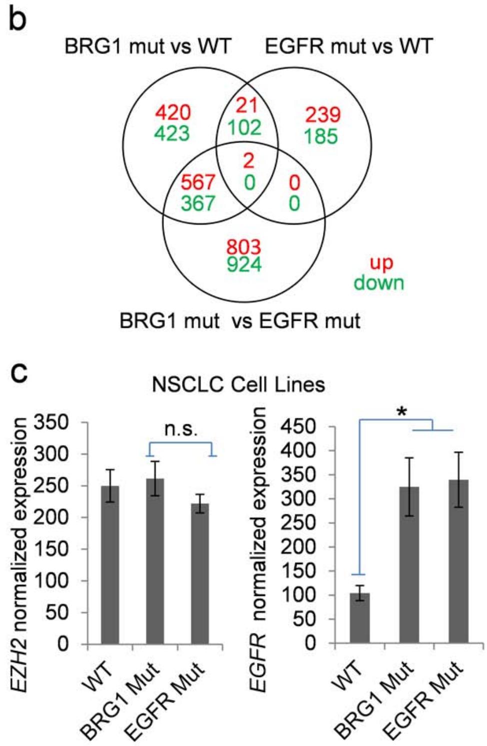
Extended Data Figure 5 | EZH2i modulates anaphase bridging differentially by genotype. **a**, Representative images of nuclei undergoing a normal anaphase and of nuclei that scored positively for the presence of anaphase bridges. **b**, Average percentage of anaphase bridging ± s.e.m. in additional *BRG1* WT H2009 and H441, *BRG1* mutant H522, and *EGFR* mutant H1650 cell lines ($n = 3$ biological replicates for all except H1650 ($n = 4$); $^*P < 0.05$). **c**, Immunofluorescence on PC9 cultures showing increase in BRG1 staining in interphase nuclei in response to EZH2i while anaphase nuclei retain strong EGFR stain, representative of three biological replicates; scale bar, 30 μm.



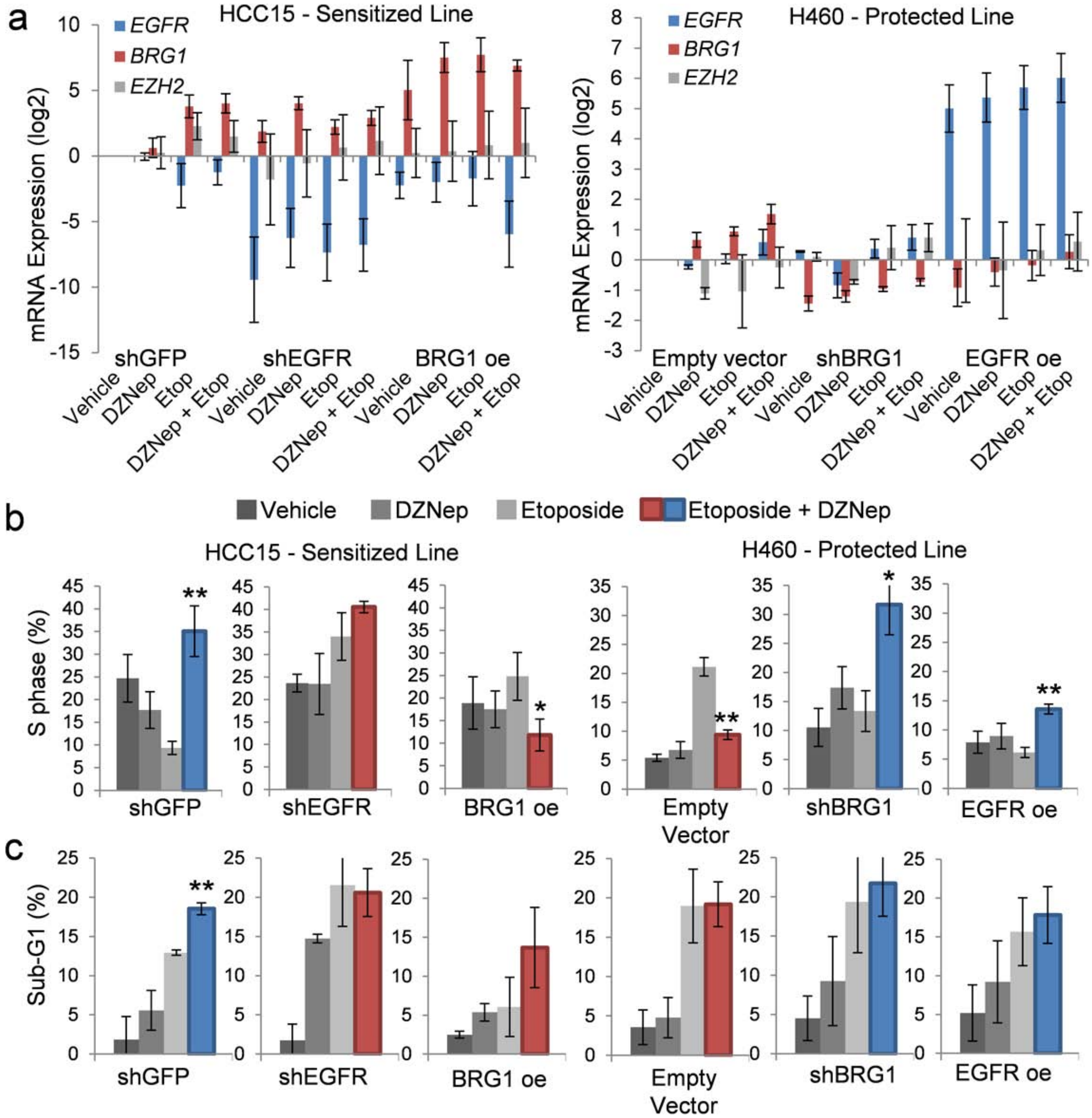
Extended Data Figure 6 | Cell cycle and apoptosis analysis of dual EZH2i- and TopoIIi-treated lines. **a**, 7AAD cell cycle flow cytometry on cultures corresponding to each experiment shown in Fig. 3d. The average percentage S phase ± s.e.m. of each culture is plotted ($n = 3$ biological replicates for H460, H23, Calu6, PC9, HCC15, A549 and H522; $n = 4$ biological replicates for Sw1573, H441 and H157).

a

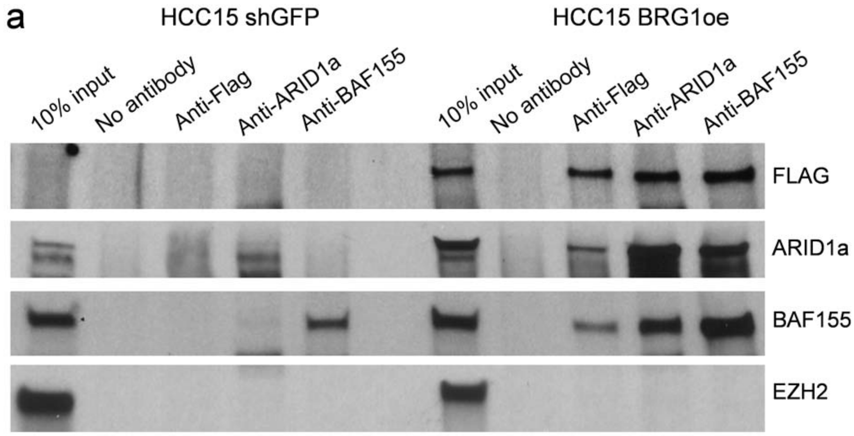
Cell Line	*EGFR*	*BRG1*
HCC2279	E746_A750del	WT
HCC2935	E746_A750del	WT
PC14	E746_A750del	WT
A427	WT	Null
H661	WT	L1161fs*3
H1703	WT	E668_Q758del
DMS-114	WT	E1310*
RERF-LC-MS	WT	A1245fs*13
H1573	WT	E1399*
H1581	WT	E1310*
H1693	WT	L1085fs*32
H838	WT	Null
H1819	WT	L1085fs*6
H596	WT	WT
H1648	WT	WT
H1437	WT	WT
H1755	WT	WT
H2087	WT	WT
H1395	WT	WT



Extended Data Figure 7 | *EGFR* and *BRG1* negatively correlate in NSCLC. **a**, Additional NSCLC cell lines with known *EGFR* and *BRG1* mutations used to estimate mutually exclusivity of the two mutations. **b**, Venn diagram of differential gene expression overlap between cell lines of various genotypes. **c**, Average probe intensity ± s.e.m. of *EGFR* probe (201983_s_at) and *EZH2* probe (203358_s_at) on the U133A Affymetrix array for cell lines with various *EGFR* and *BRG1* mutational statuses ($n = 6$ per genotype, see Methods; $^*P = 0.014$).



Extended Data Figure 8 | Modulation of EGFR and BRG1 influences sensitized and protected phenotypes. **a**, RT-qPCR for average expression of *BRG1*, *EGFR* and *EZH2* ± s.e.m. in the various indicated treated transduced cell lines ($n = 3$ biological replicates). **b**, For the indicated HCC15 and H460 stably transduced etoposide-treated cell lines, 7AAD flow cytometry was used to assess average changes in percentage S phase ± s.e.m. in response to DZNep ($n = 3$ biological replicates; $*P = 0.02$, $**P < 0.001$). **c**, Average percentage sub-G_1 fractions ± s.e.m. of the indicated 4-day cultures were assessed during 7AAD cell cycle flow cytometry analysis. Critically, for these assays, the supernatant of each culture was retained and combined with the trypsinized adherent cells to reflect the total amount of apoptosis/necrosis in each culture accurately ($n = 3$ biological replicates; $**P = 0.03$).



Extended Data Figure 9 | Confirmation that BRG1 re-expression leads to formation of BAF complex. **a**, Immunoprecipitation of BAF complex members from nuclear lysates of the (left) *BRG1* mutant HCC15 shGFP control cell line and (right) the HCC15 line with BRG1 re-expressed shows that exogenously expressed Flag-tagged BRG1 does result in BRG1-containing BAF complex formation. The blot is representative of three biological replicates.

LETTER

doi:10.1038/nature14176

Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation

Tuncay Baubec[1], Daniele F. Colombo[1], Christiane Wirbelauer[1], Juliane Schmidt[1], Lukas Burger[1,2], Arnaud R. Krebs[1], Altuna Akalin[1]† & Dirk Schübeler[1,3]

DNA methylation is an epigenetic modification associated with transcriptional repression of promoters and is essential for mammalian development. Establishment of DNA methylation is mediated by the *de novo* DNA methyltransferases DNMT3A and DNMT3B, whereas DNMT1 ensures maintenance of methylation through replication[1]. Absence of these enzymes is lethal[2], and somatic mutations in these genes have been associated with several human diseases[3,4]. How genomic DNA methylation patterns are regulated remains poorly understood, as the mechanisms that guide recruitment and activity of DNMTs *in vivo* are largely unknown. To gain insights into this matter we determined genomic binding and site-specific activity of the mammalian *de novo* DNA methyltransferases DNMT3A and DNMT3B. We show that both enzymes localize to methylated, CpG-dense regions in mouse stem cells, yet are excluded from active promoters and enhancers. By specifically measuring sites of *de novo* methylation, we observe that enzymatic activity reflects binding. *De novo* methylation increases with CpG density, yet is excluded from nucleosomes. Notably, we observed selective binding of DNMT3B to the bodies of transcribed genes, which leads to their preferential methylation. This targeting to transcribed sequences requires SETD2-mediated methylation of lysine 36 on histone H3 and a functional PWWP domain of DNMT3B. Together these findings reveal how sequence and chromatin cues guide *de novo* methyltransferase activity to ensure methylome integrity.

Several models of DNMT targeting have been proposed based on biochemical, structural and functional studies[5]. These include readout of histone modifications, binding to nucleosomes, recruitment by accessory proteins, as well as RNA-mediated or DNA sequence-specific targeting or repulsion[5]. To gain insights into targeting *in vivo*, we determined the genomic binding of DNMT3A and DNMT3B in mouse embryonic stem (ES) cells. We used chromatin immunoprecipitation (ChIP) of biotin-tagged proteins, which are expressed from a defined genomic site as previously described[6] (Extended Data Fig. 1a). Here we expressed biotin-tagged versions of the major catalytically active DNMT3 isoforms present in mouse ES cells: DNMT3A2 and DNMT3B2[7] (Extended Data Fig. 1b, c), and used biotin-based ChIP to generate genome-wide binding maps (Extended Data Fig. 1d–g). This revealed a broad genomic binding pattern (Fig. 1a) with preferential localization to methylated CpG-rich regions (Fig. 1b and Extended Data Fig. 2a), a finding that links both enzymes to sites of DNA methylation. Furthermore, both paralogues did not bind to active regulatory sites that display low levels of DNA tri-methylation[8], such as distal regulatory elements (LMR) and unmethylated regions (UMR) including H3K4me3 positive CpG islands (Fig. 1b). This absence of binding is compatible with repulsion by H3K4 methylation, as previously suggested for ADD domain proteins[9].

Besides these similarities, we also observed notable differences in binding between DNMT3A2 and DNMT3B1 (Fig. 1a). Analysis of enriched sites indicates that DNMT3B1-bound sites primarily consist of gene bodies (Extended Data Fig. 2b). We compared DNMT3A and DNMT3B1 binding with the genomic distributions of various histone marks, 5-hydroxymethylcytosine (5-hmC), as well as transcription and RNA polymerase II (Fig. 1c and Extended Data Fig. 2c). This revealed that all features of active transcription such as mRNA, RNA pol II and H3K36me3 were exclusively enriched for DNMT3B1-bound sites (Fig. 1c and Extended Data Fig. 2c). Similar results were obtained by correlation of 1-kb windows over the entire genome (Extended Data Fig. 2d, e). Thus besides the general binding to methylated CpG-rich regions, DNMT3B1 exhibits additional preferences for actively transcribed genes in mouse stem cells (Fig. 1d).

Before further exploring this potential link to active genes, we first asked how localization relates to enzymatic activity. More specifically, we re-introduced DNMT3A2 or DNMT3B1 individually into stem cells that lack DNA methylation due to genetic deletions of all *Dnmt* genes (*Dnmt* triple knockout (TKO))[10] (Extended Data Fig. 3a). In this experimental setting, any measured methylation results from the transfected *de novo* DNMT. Genomic binding of the reintroduced DNMTs in TKO cells resembles that in wild-type ES cells (Extended Data Fig. 3b), indicating that recruitment is independent of pre-existing DNA methylation. Moreover we observe reproducible rates of re-methylation measured by mass spectrometry, even in the absence of the maintenance methyltransferase DNMT1. Here DNMT3A2 showed higher overall activity compared to DNMT3B1 with 0.7% versus 0.18% methylation for all cytosines, respectively, compared to the 4.2% observed in wild-type stem cells (Extended Data Fig. 3c). To locate chromosomal sites that undergo *de novo* methylation, we generated bisulphite maps from the same lines. These confirmed the mass spectrometry measurements, as they reveal a total of 7% methylation for DNMT3A2 and 2.8% for DNMT3B1 at cytosines in a CpG dinucleotide context (Extended Data Fig. 3d). In both cases, *de novo* methylation did not occur at UMRs and was reduced at LMRs (Fig. 2a) in agreement with the observed absence of binding. In contrast, sites bound by DNMT3A2 or DNMT3B1 displayed increased *de novo* methylation in TKO cells (Fig. 2a and Extended Data Figs 3e and 4a). Interestingly, *de novo* methylation occurs more often at adjacent CpGs, and is furthermore distance–dependent even at the level of single molecules (Fig. 2b and Extended Data Fig. 3f, g). This indicates a processive behaviour, which appears more pronounced for DNMT3A2 than for DNMT3B1. This could explain the preferential localization of DNMTs to CpG-dense regions in both wild-type and TKO cells (Fig. 1b and Extended Data Fig. 2a). The observed distance–dependence deviates from a monotonically decreasing function at around 180 base pairs (bp), where it reaches a local maximum (Fig. 2b and Extended Data Fig. 3f). As this resembles the size of one nucleosomal unit, we wanted to test if this periodicity reflects nucleosomal organization. We focused our analysis on positioned nucleosomes that occur adjacent to binding sites of the insulator protein CTCF (Extended Data Fig. 3h). Indeed, *de novo* methylation by both paralogues is occluded from the core of these nucleosomes but occurs predominantly in their linker regions (Fig. 2c and Extended Data Fig. 3i). This *in vivo* activity

[1]Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland. [2]Swiss Institute of Bioinformatics. Maulbeerstrasse 66, CH-4058 Basel, Switzerland. [3]University of Basel, Faculty of Sciences, Petersplatz 1, CH-4001 Basel, Switzerland. †Present address: Max Delbrück Center, The Berlin Institute for Medical Systems Biology, Robert Rössle Strasse 10, DE-13125 Berlin, Germany.

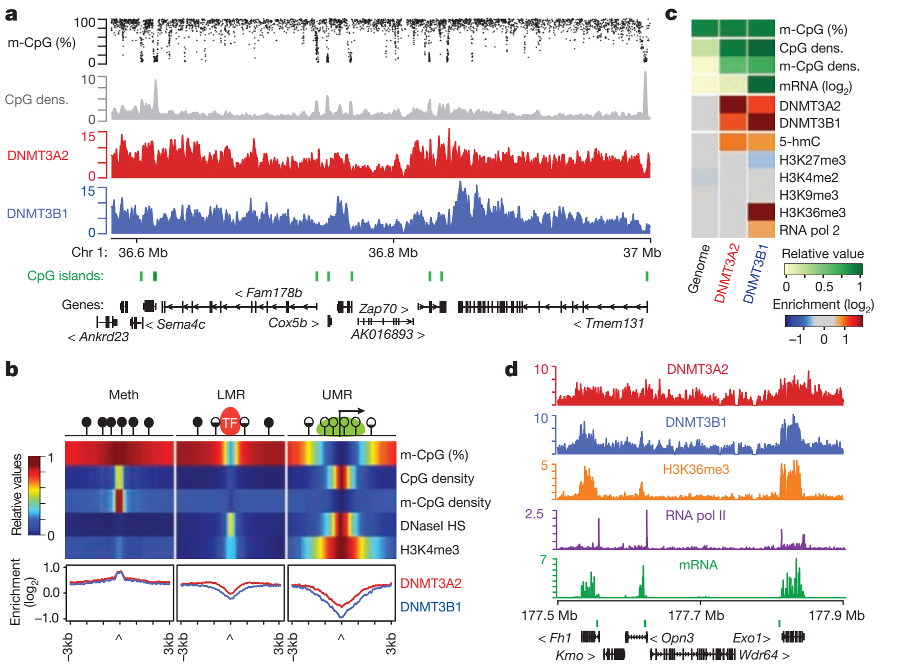


Figure 1 | DNMT3A2 and DNMT3B1 show similar localization to most parts of the genome, but are differently recruited to gene bodies. **a**, DNMT protein binding to a representative genomic region is shown as number of reads per 100 base pairs (bp). The top tracks display frequency of CpG methylation (m-CpG %) and CpG density per 100 bp. Location of CpG islands (UCSC) and genes (RefSeq) are displayed below. **b**, Average density plots of DNMT3 binding at methylated CpG-dense regions (meth, $n = 3{,}626$), low methylated regions (LMR) bound by transcription factors ($n = 8{,}249$) and unmethylated regions (UMR, $n = 3{,}117$). Heatmap representation indicates genomic and epigenomic features at these sites and average DNMT enrichment is shown below. **c**, Features at DNMT3-bound sites (DNMT3A2, $n = 6{,}499$; DNMT3B1, $n = 4{,}799$) in comparison to genome average. **d**, Representative genomic region to illustrate DNMT3B1 binding to transcriptionally active, H3K36me3-positive genes. Track displays number of normalized sequencing reads per 100 bp.

measure is compatible with the reported enzymatic behaviour *in vitro*[11] and genome-wide methylation patterns[8,12], and supports a model in which nucleosome positioning can occur upstream of *de novo* methylation.

The above results indicate that *de novo* methylation is primarily regulated at the level of recruitment, while positioned nucleosomes may further direct this process. In the case of DNMT3B1, *de novo* methylation is preferentially targeted to genomic regions with elevated H3K36me3 levels (Fig. 2d and Extended Data Fig. 4a), reflecting the DNMT3B1 recruitment to regions of active transcription. Further supporting this observation, DNMT3B1-dependent *de novo* DNA methylation in TKO cells appears to be enhanced at the bodies of highly transcribed genes, whereas methylation set by DNMT3A2 shows an opposite behaviour, with decreased methylation of highly transcribed genes (Extended Data Fig. 4b). This methylation activity of DNMT3B furthermore mirrors enzyme binding at gene bodies (Extended Data Fig. 4c).

Having established that local binding of DNMT3B1 to sites of active transcription translates into *de novo* methylation, we next asked if this contributes to DNA methylation maintenance, even in the presence of

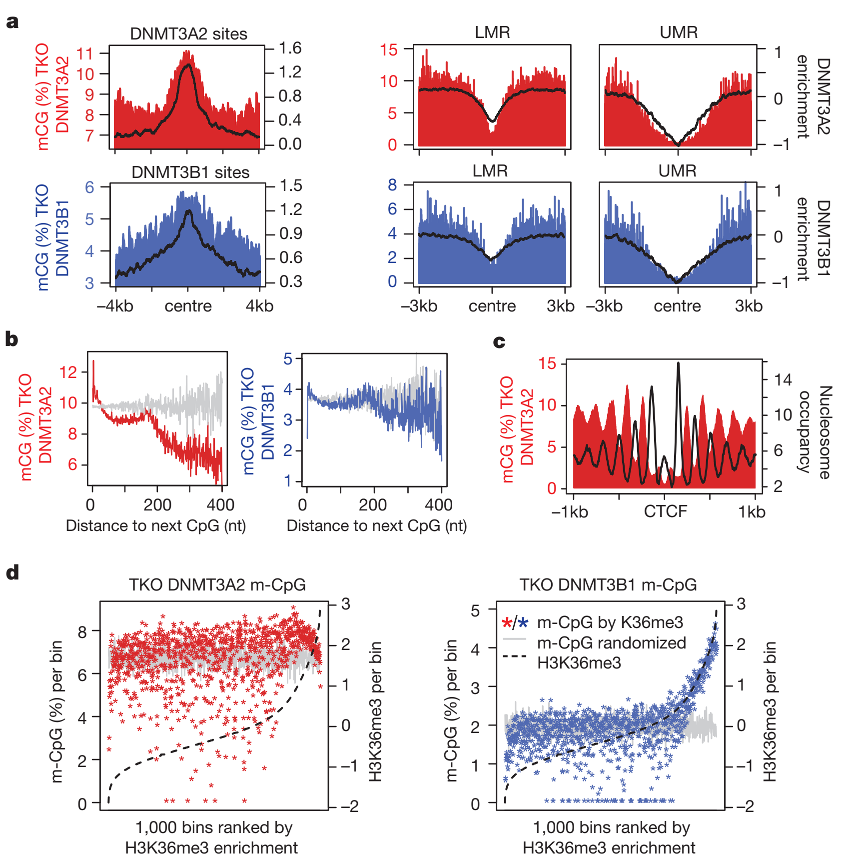


Figure 2 | *De novo* methylation relative to DNMT binding and chromatin features. **a**, DNMT3A2 and DNMT3B1-bound sites are preferentially *de novo* methylated. Shown is average methylation (red, blue) and DNMT binding (black line) for bound sites and regulatory regions (LMR/UMR) after reintroducing each enzyme into TKO cells. **b**, *De novo* methylation follows a periodicity similar to nucleosomal length. Shown is the average methylation (y axis) based on distance (nt, nucleotides) to the nearest CpG (x axis). Measured values are shown in red or blue, randomized values in grey. **c**, *De novo* methylation occurs preferentially at linker regions between nucleosomes. Shown is DNA methylation and nucleosome positioning (MNaseI, black) around genomic binding sites of CTCF ($n = 4{,}126$). **d**, Preferential *de novo* methylation of H3K36 tri-methylated sites by DNMT3B1. Asterisks display *de novo* DNA methylation genome-wide in relation to H3K36me3 enrichment. 1-kb-sized genomic tiles were ranked and grouped by H3K36me3 enrichment (1,000 bins, 433 tiles per bin). Red and blue indicate median DNA methylation per bin in TKO cells expressing DNMT3A2 or DNMT3B1, respectively. Dashed line indicates H3K36me3 enrichment per bin. DNA methylation after random grouping of genomic tiles is shown as control in grey.

DNMT1. Towards this goal, we performed parallel, targeted bisulphite sequencing analysis in wild-type and *Dnmt3a* or *Dnmt3b* single knock-out (KO) ES cells, respectively[2] (Extended Data Fig. 5a, b and Methods). This revealed that the absence of either DNMT3A or DNMT3B did not affect DNA methylation at sites bound by both enzymes (Extended Data Fig. 5c, d), indicating redundancy at these sites as reported previously[2,7]. However, DNA methylation was significantly reduced at several DNMT3B-specific targets in cells that lack DNMT3B (Extended Data Fig. 5c, d). These sites are actively transcribed and H3K36me3-positive, indicating a role for DNMT3B in regulating DNA methylation at active genes in stem cells (Extended Data Fig. 5e, f).

The genome-wide binding analysis suggested a link between transcription and DNMT3B1 activity. DNMT3B1 binding to genes, but not DNMT3A2, follows transcriptional activity (Extended Data Fig. 4c). This dependence on transcription and H3K36me3 is similar between exons and introns, even though exons are CpG rich and introns are CpG poor (Fig. 3a and Extended Data Fig. 6a–c). This link is also evident when applying a linear model to predict DNMT3B1 binding using various genomic and epigenomic features. Again, methyl-CpG density partially predicts binding for both DNMT3A2 and DNMT3B1, whereas H3K36me3 predicts only DNMT3B1 binding (Extended Data Fig. 6d). H3K36 tri-methylation is deposited by SETD2 at active genes but only downstream of the promoter, where it binds the elongation-competent and serine 2 phosphorylated form of RNA polymerase II (ref. 13). As a result, the level of this histone modification increases with the rate of transcription[14] (Fig. 3b and Extended Data Fig. 6e). DNMT3B1 binding to gene bodies similarly scales with H3K36me3 levels in a transcription-dependent manner and occurs also more prominently downstream of the transcription start site (TSS) (Fig. 3b). In agreement, DNMT3B1 binding and *de novo* methylation at transcribed introns was only observed for instances positive for H3K36me3, but not at transcribed sites lacking H3K36me3 (Extended Data Fig. 6f-h), suggesting that H3K36me3, rather than transcription, could directly influence DNMT3B recruitment.

To test if changes in H3K36me3 distribution result in DNMT3B1 relocalization, we differentiated ES cells constitutively expressing DNMT3B1 to neuronal progenitor (NP) cells. Indeed, we observe that changes in DNMT3B1 binding to gene bodies positively correlate with H3K36me3 changes (Fig. 3c, d and Extended Data Fig. 6i), revealing cell-type-specific binding of DNMT3B1 to transcribed genes.

To test if ongoing transcription is required for DNMT3B1 binding to gene bodies, we chemically inhibited elongation of RNA pol II for 10 h by 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) treatment, which does not immediately affect H3K36me3 levels (Extended Data Fig. 7a). Even in the generated absence of RNA polymerase II activity, DNMT3B1 continues to bind to H3K36me3-positive gene bodies (Extended Data Fig. 7b, k, l). This argues against a scenario in which DNMT3B1 interacts with components of the elongating polymerase or is attracted by increased DNA accessibility at sites of polymerase passage. Instead, it emphasizes a requirement of H3K36me3. We tested this experimentally through generation of ES cells that lack *Setd2* using the CRISPR/Cas9 system, resulting in depletion of H3K36me3 (Extended Data Fig. 7c–f). The loss of H3K36me3 impairs recruitment of DNMT3B1 to active genes, while binding of DNMT3A2 is unaffected (Fig. 4a, b and Extended Data Fig. 7g–l and 8a). At the same time, binding of DNMT3B1 to CpG-rich methylated sites remains unchanged, resulting in a genomic binding pattern that closely resembles that of DNMT3A2 (Fig. 4a, b and Extended Data Fig. 7h–j and l). To determine if H3K36me3 is also required for preferential *de novo* methylation by DNMT3B, we also deleted *Setd2* in *Dnmt* TKO cells with reintroduced DNMT3B1. This did not affect *de novo* methylation of CpG-rich sequences and linker regions but particularly depleted DNA methylation at H3K36me3 sites (Fig. 4c and Extended Data Fig. 8b–f). Thus SETD2-mediated H3K36me3 can guide DNMT3B binding and *de novo* DNA methylation to transcribed gene bodies in mouse stem cells.

Preceding work suggests that PWWP domains that are also present in DNMT3A and DNMT3B could be directly involved in H3K36me3 readout[15,16] (Extended Data Fig. 9a–d). *In vitro* interaction experiments indeed indicate that the amino-terminal part of DNMT3B (amino acids 1 to 427) preferentially interacts with H3K36 methylated nucleosomes (Extended Data Fig. 9e, f). These interactions are disrupted by point mutations in the PWWP domain, including one previously identified

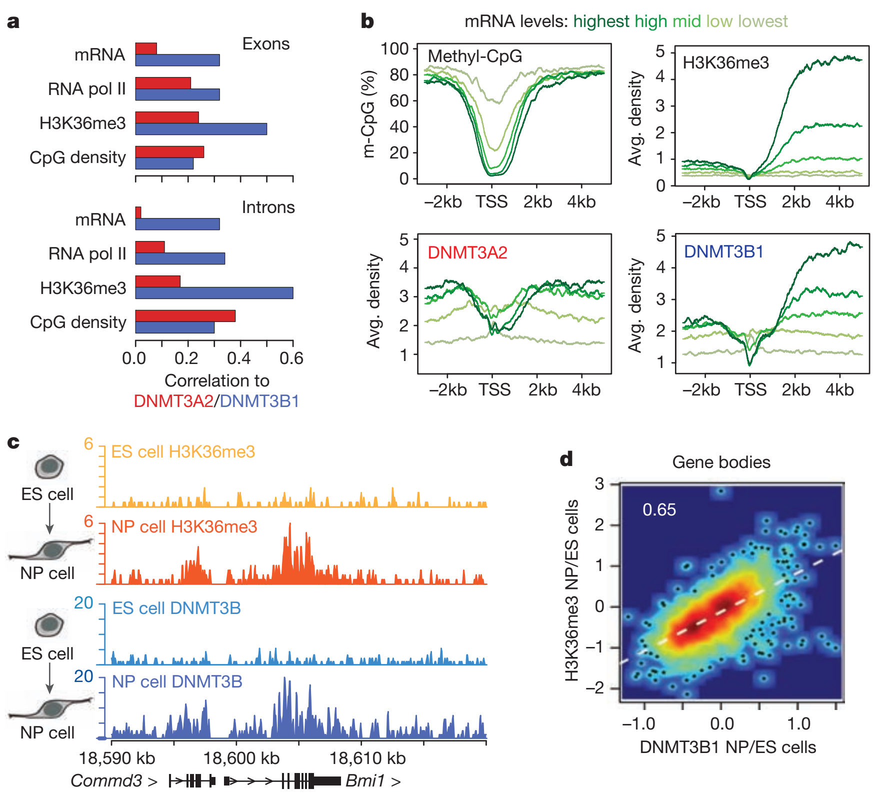


Figure 3 | DNMT3B binding follows co-transcriptional deposition of H3K36me3. **a**, DNMT3B1 presence at exons and introns correlates with transcription. Pearson's correlation coefficient between DNMT binding and features of exons and introns is shown. **b**, Average profiles around genes, grouped based on transcriptional activity (2,051 genes per bin). **c**, Representative gene locus that becomes transcriptionally active during differentiation of embryonic stem (ES) cells to neuronal progenitor (NP) cells and gains H3K36me3 and DNMT3B1 binding. **d**, Dynamics of H3K36me3 deposition correlate with dynamics in DNMT3B1 binding during this neuronal differentiation. Shown are the differences in DNMT3B and H3K36me3 signal of the two stages measured at gene bodies bound by DNMT3B1 ($n = 1{,}944$). Pearson's correlation coefficient is shown.

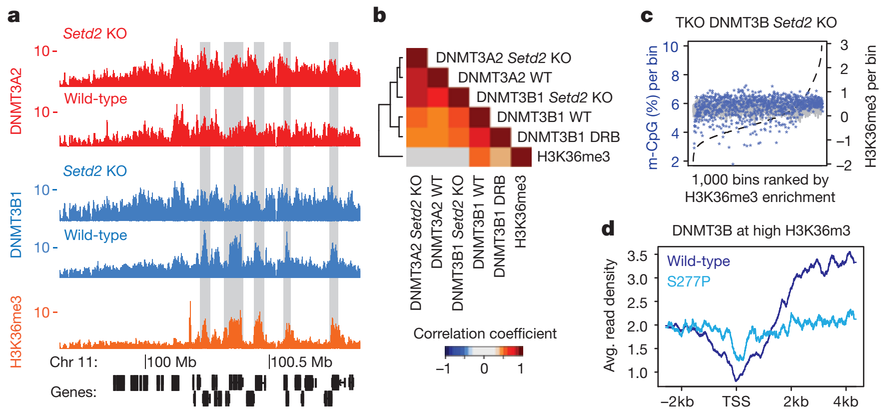


Figure 4 | The H3K36 methyltransferase SETD2 and the PWWP domain of DNMT3B are required for targeting *de novo* methylation to transcribed genes. **a**, DNMT3B1 shows reduced binding to active genes in the absence of SETD2 and H3K36me3, whereas DNMT3A2 remains unchanged. Active genes are indicated in grey. **b**, Unsupervised clustering of correlation coefficients reveals global loss of DNMT3B1 binding to H3K36me3-positive gene bodies in the absence of SETD2 (10,257 gene bodies). 'DRB' indicates DNMT3B1 binding following 10 h inhibition of RNA polymerase elongation by DRB. **c**, Absence of SETD2 leads to lack of *de novo* methylation at sites that have high H3K36me3 sites in wild-type cells (see Fig. 2d for comparison). **d**, An ICF syndrome-related mutation (S277P) in the PWWP domain leads to loss of binding at active gene bodies. Shown are average profiles for wild-type DNMT3B1 (dark blue) and S277P mutation (light blue) at highly transcribed genes in ES cells marked by H3K36me3 ($n = 2{,}051$).

in ICF (immunodeficiency, centromere instability and facial anomalies) syndrome patients[17] (human, S282P; mouse, S277P). In ES cells, the same mutations in the PWWP domain result in loss of interactions with H3K36me3-positive chromatin and recruitment to active gene bodies (Fig. 4d and Extended Data Fig. 10a–e). Thus, the PWWP domain appears necessary for genomic targeting of DNMT3B1 to chromatin marked by H3K36 tri-methylation, supporting a model of direct interaction. However, since PWWP domains have been also shown to mediate interactions with DNA and other histone modifications[18,19], this model awaits further biochemical characterization.

This study suggests that the mammalian *de novo* DNMTs display complex genomic binding that translates into site-specific *de novo* methylation activity in mouse stem cells. Preferential binding and enzymatic activity at regions with increased CpG density, outside of active regulatory elements or positioned nucleosomes, suggest that substrate availability and accessibility are global determinants of *de novo* DNA methylation. Of particular relevance is our observation of DNMT3B1 recruitment and *de novo* methylation to active genes that scales with co-transcriptional deposition of H3K36me3. Genetic ablation of SETD2, the enzyme that sets H3K36me3, disrupts DNMT3B1 binding and *de novo* methylation at H3K36me3 sites, suggesting a direct requirement for this mark. In addition, substitutions in the putative H3K36me3 reader domain PWWP abolish the association of DNMT3B1 with chromatin and its preference to sites of active transcription. Notably, our results do not account for the differences in binding between DNMT3A2 and DNMT3B, which are highly similar in domain composition. At this point we can only speculate that this differential binding could be regulated by protein–protein interactions or post-translational modifications. For example, methylation of transcribed genes is highly prominent in oocytes yet depends on DNMT3A and DNMT3L[20], suggesting that cellular context could define the responsible enzyme for gene body methylation. Preferential binding of DNMT3B appears functionally relevant for DNA methylation of H3K36me3-positive gene bodies in stem cells. In line with our findings, DNMT3B is associated with H3K36me3 in human embryonic carcinoma cells[21] and methylation of gene bodies in human colon cancer cell lines was recently shown to be dependent on DNMT3B and active transcription[22]. Transcribed gene bodies display increased levels of 5-hmC[23], suggesting that targeting of DNMT activity might compensate for active demethylation. In addition, genic methylation patterns appear to substantially vary in mature cardiomyocytes[24] and medulloblastomas[25] arguing for tissue- and disease-specific targeting and maintenance of DNA methylation.

Gene bodies are a widely conserved target of methylation in eukaryotic cells. This is observed in the majority of organisms with mosaic DNA methylation, ranging from anemones to plants to insects, and is associated with active transcription[26–28]. It thus has been suggested that gene body methylation could function similar to H3K36 tri-methylation to promote proper splicing and compact chromatin at active genes[29,30]. Our results show that regulated targeting of DNA methylation to transcribed gene bodies can still be detected in organisms with genome-wide methylation such as mammals. It is thus tempting to speculate that the proposed model of methylation recruitment via co-transcriptional modification of histones could generally account for genic DNA methylation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 May; accepted 23 December 2014.
Published online 21 January 2015.

1. Goll, M. G. & Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74,** 481–514 (2005).
2. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99,** 247–257 (1999).
3. Xu, G. L. *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402,** 187–191 (1999).
4. Yan, X.-J. *et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene *DNMT3A* in acute monocytic leukemia. *Nature Genet.* **43,** 309–315 (2011).
5. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nature Rev. Genet.* **14,** 204–220 (2013).
6. Baubec, T., Ivanek, R., Lienert, F. & Schübeler, D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153,** 480–492 (2013).
7. Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23,** 5594–5605 (2003).
8. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480,** 490–495 (2011).
9. Ooi, S. K. T. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **448,** 714–717 (2007).
10. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11,** 805–814 (2006).
11. Felle, M. *et al.* Nucleosomes protect DNA from DNA methylation *in vivo* and *in vitro*. *Nucleic Acids Res.* **39,** 6956–6969 (2011).
12. Kelly, T. K. *et al.* Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22,** 2497–2506 (2012).

13. Krogan, N. J. *et al.* Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.* **23,** 4207–4218 (2003).
14. Tippmann, S. C. *et al.* Chromatin measurements reveal contributions of synthesis and decay to steady-state mRNA levels. *Mol. Syst. Biol.* **8,** 593 (2012).
15. Dhayalan, A. *et al.* The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* **285,** 26114–26120 (2010).
16. Kungulovski, G. *et al.* Application of histone modification-specific interaction domains as an alternative to antibodies. *Genome Res.* **24,** 1842–1853 (2014).
17. Shirohzu, H. *et al.* Three novel *DNMT3B* mutations in Japanese patients with ICF syndrome. *Am. J. Med. Genet.* **112,** 31–37 (2002).
18. Qiu, C., Sawada, K., Zhang, X. & Cheng, X. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nature Struct. Biol.* **9,** 217–224 (2002).
19. Qin, S. & Min, J. Structure and function of the nucleosome-binding PWWP domain. *Trends Biochem. Sci.* **39,** 536–547 (2014).
20. Smallwood, S. A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature Genet.* **43,** 811–814 (2011).
21. Jin, B. *et al.* Linking DNA methyltransferases to epigenetic marks and nucleosome structure genome-wide in human tumor cells. *Cell Reports* **2,** 1411–1424 (2012).
22. Yang, X. *et al.* Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **26,** 577–590 (2014).
23. Pastor, W. A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473,** 394–397 (2011).
24. Gilsbach, R. *et al.* Dynamic DNA methylation orchestrates cardiomyocyte development, maturation and disease. *Nature Commun.* **5,** 5288 (2014).
25. Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510,** 537–541 (2014).
26. Tweedie, S., Charlton, J., Clark, V. & Bird, A. Methylation of genomes and genes at the invertebrate–vertebrate boundary. *Mol. Cell. Biol.* **17,** 1469–1475 (1997).
27. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328,** 916–919 (2010).
28. Nanty, L. *et al.* Comparative methylomics reveals gene-body H3K36me3 in *Drosophila* predicts DNA methylation and CpG landscapes in other invertebrates. *Genome Res.* **21,** 1841–1850 (2011).
29. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nature Rev. Genet.* **9,** 465–476 (2008).
30. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Rev. Genet.* **13,** 484–492 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to S. Dessus-Babus, K. Jacobeit, T. Roloff (FMI), I. Nissen, and C. Beisel (ETH BSSE) for processing deep-sequencing samples. We thank R. Sack for processing histone H3K27/36 methylation measurements and analysis. We further thank L. Hoerner for technical assistance and M. Stadler and D. Gaidatzis for bioinformatic advice and providing tools to process and analyse deep-sequencing data. FMI imaging and FACS facilities for technical support. P. Ginno (FMI), N. Thomae (FMI), and M. Lorincz (UBC Vancouver) for advice and critical comments on the manuscript. Research in the laboratory of D.S. is supported by the Novartis Research Foundation, the European Union (NoE ‘‘EpiGeneSys’’ FP7- HEALTH-2010-257082 and the ‘‘Blueprint’’ consortium FP7-282510), the European Research Council (EpiGePlas), the SNF Sinergia program, and the Swiss initiative in Systems Biology (RTD Cell Plasticity). J.S. acknowledges support from the Boehringer Ingelheim Fonds. T.B. and A.R.K. acknowledge support from EMBO postdoctoral long-term fellowships.

Author Contributions T.B. initiated the study; T.B. and D.S. designed experiments; cell lines were generated by T.B. and D.F.C.; genomics experiments were performed by T.B.; interaction experiments were performed by T.B. and C.W.; genome-wide binding and methylation data was analysed by T.B. and L.B.; T.B., A.R.K., J.S., A.A. and L.B. designed/provided experimental and analytical tools; the manuscript was prepared by T.B. and D.S. All authors discussed the results and commented on the manuscript.

Author Information Genome-wide datasets generated for this study are deposited at GEO under the accession number GSE57413. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.B. (tuncay.baubec@fmi.ch) or D.S. (dirk@fmi.ch).

METHODS

Cell lines and cell culture. Mouse embryonic stem cells were generated, maintained and differentiated as described previously[6]. HA36CB1 and DNMT TKO-133 cells were cultivated on 0.2% gelatine-coated dishes in DMEM (Invitrogen) supplemented with 15% fetal calf serum (Invitrogen), 1× non-essential amino acids (Invitrogen), 1 mM L-glutamine, LIF and 0.001% β-mercaptoethanol. Differentiation to neuronal progenitors was performed as previously described[31]. cDNAs encoding DNMT3A2 and DNMT3B1 were amplified from ES cell mRNA extracts and cloned into pL1-CAGGS-bio-MCS-polyA-1L, pL1-CMV-bio-MCS-polyA-1L or pCAGGS-bio-MCS-IRES-BlasticidinR-polyA. Amino acid substitutions were introduced via quick-change PCR. Detailed cloning strategies and constructs are available upon request. Biotin-tagged DNMT3A2 and DNMT3B1 wild-type and *Setd2* KO ES cell lines were obtained by recombinase-mediated cassette exchange (RMCE) as previously described[6]. For 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) treatments, 75 μM DRB (diluted in DMSO) were added directly to the cell culture medium and incubated for 10 h. CRISPR/Cas9 deletions of SETD2 were generated using pX330 (ref. 32) and guide design was performed using the CRISPR design tool[33]. The *Setd2* deletion was targeted to exon 3. Guide sequences are available upon request.

Biotin–streptavidin chromatin immunoprecipitation. Chromatin immunoprecipitation (ChIP) was performed as described[6]. In brief, $10–20 \times 10^6$ cells were fixed for 10 min with 1% formaldehyde, followed by chromatin extraction and sonication. Immunoprecipitation was performed overnight at 4 °C using pre-blocked (0.1% cold fish skin gelatin and 100 ng yeast tRNA) streptavidin-M280 magnetic beads on 150–200 μg pre-cleared chromatin. ChIP washes were performed by two rounds of 2% SDS, 1× high-salt buffer, 1× LiCl buffer, and two rounds of TE. Elution was performed by RNaseA treatment for 30 min at 37 °C and proteinase K treatment for 3 h at 65 °C, with subsequent decrosslinking overnight at 55 °C. DNA was purified by phenol/chloroform extraction and ethanol precipitation.

For biotin–streptavidin ChIP followed by immunoblot analysis, we performed standard ChIP as above, with some modifications. We replaced the TE washes with two washes in TBS 0.1% Tween-20, resuspended the beads in Laemmli buffer and boiled for 15 min to decrosslink the proteins. Enriched and input chromatin was analysed by SDS–PAGE and western blotting to PVDF membranes using antibodies specific to H3K36me3 (ab9050) or total H3 (ab1791).

Biotin–streptavidin immunodetection and cellular fractionation. Whole-cell extracts were obtained from ES cells by resuspension in 400 mM NaCl, 0.3% NP40, 5 mM EDTA and 50 mM TRIS pH 7.5 followed by 10 min incubation at 4 °C. Equal amounts of protein were resuspended in Laemmli buffer before SDS–PAGE and transfer to PVDF membranes. Cytoplasmic extracts were directly obtained by resuspending ES cells in 20 mM HEPES, 10 mM KCl, 10% glycerol, 2.5 mM $MgCl_2$ and 0.2% Tween followed by centrifugation at 1,000*g*. Supernatant contained the cytoplasmic fraction. Pellets containing nuclei were washed once in cytoplasmic extract buffer, then incubated in 10 mM TRIS-HCl (pH 7.5), 60 mM KCl, 100 mM NaCl, 1.5 mM MgCl, 1 mM $CaCl_2$, 250 mM sucrose, 10% glycerol, 0.15% NP40 and 50 U DNaseI (Takara) for 20 min at room temperature, then 10 min at 4 °C, followed by centrifugation at 600*g*. Supernatant contained the nuclear soluble fraction. The pellet containing the chromatin fraction was subsequently extracted by repeating the above step using 500 mM NaCl. Equal amounts of protein from all three fractions were resuspended in Laemmli buffer before SDS–PAGE and western blot transfer to PVDF membranes. Membranes were blocked with 5% milk or 5% BSA. Used antibodies were, DNMT3A (IMG-268A), DNMT3B (IMG-184A), LAMIN-B1 (sc6216), Pol2-S2P (ab24758) and streptavidin-HRP (Invitrogen).

Mononucleosome interaction assays. GST-fused, N-terminal DNMT3B region (amino acids 1–427) and PWWP domain (amino acids 227–343) were expressed using pGEX-4T1 in *E. coli*. GST-fusion proteins were pre-immobilized on 50 μl glutathione sepharose 4B beads (at saturating protein concentrations). A 4 μg sample of HeLa mononucleosomes (Epicypher) was incubated with 50 μl immobilized proteins (50% slurry) in 200 μl mononucleosome interaction buffer (100 mM KCl, 5% glycerol, 10 mM Tris HCl pH 8, protease inhibitors and 1 mM DTT) overnight at 4 °C. Resin was washed with 1ml interaction buffer three times, with tube changes. Resin was resuspended in Laemmli buffer, boiled and bound histones were separated on 15% SDS–PAGE followed by transfer to PVDF membranes. Detection was performed using the following antibodies: H3K36me3 (ab9050), total H3 (ab1791), H3K27me3 (07-449), H2A119ub (05-678). 2% of resin was used to indicate loading of GST-fused proteins, where detection was performed either using anti-GST antibodies (NEB) or Coomassie staining.

Immunofluorescence. For immunofluorescence, cells were directly cultivated on poly-L-lysine-coated coverslips, fixed for 10 min in 3% PFA and permeabilized in 0.1% NaCitrate and 0.1% Triton X-100. After 30 min blocking with 0.1% Tween-20, 3% BSA (w/v) and 10% normal goat serum in PBS, detection was performed with Streptavidin-AF568 (Invitrogen) or anti-DNMT3B (IMG-184A) overnight at 4 °C. Coverslips were washed with 0.1% Tween-20 and 0.25% BSA (w/v) in PBS. Nuclei were counterstained with DAPI (4′,6-diamidino-2-phenylindole). Images were acquired using a Zeiss Z1 epifluorescence microscope. Image analysis and preparation was done with Zeus (Zeiss) and Photoshop (Adobe).

ChIP-seq library preparation and high-throughput sequencing. For ChIP-seq, sequencing libraries were prepared using the NEB-next ChIP-seq library Kit (E62402L) following standard protocols. Four to five samples with different index barcodes were combined at equal molar ratios and sequenced as pools. Sequencing of library pools was performed on Illumina HiSeq 2000 machines according to Illumina standards, with 50 bp single-end sequencing. Library de-multiplexing was performed following Illumina standards.

Whole-genome bisulphite library preparation and sequencing. Whole-genome bisulphite sequencing was performed as described[8], with slight modifications. A 6 μg sample of sonicated, genomic DNA (bulk = 300 bp), including spiked-in conversion controls of *Sss*I (NEB)-methylated phage T7 and unmethylated phage Lambda DNA (both 10 ng, sonicated), was end repaired and 3′-end adenylated using Illumina paired-end sample preparation kit (Illumina, 1001809). Methylated paired-end library adapters (Illumina, 1006132) were ligated to the DNA fragments as per manufacturer's instructions for genomic DNA library construction. Adaptor-ligated DNA was isolated by 2% agarose gel electrophoresis (350–400 bp). Gel-purified DNA was then converted by sodium bisulphite using the Qiagen Epitect bisulphite conversion kit with slight modifications according to Illumina whole-genome bisulphite sequencing library protocol. Converted libraries were enriched by 10 cycles of PCR with the following reaction composition: 2.5 U PfuTurboCx Hotstart DNA polymerase (Stratagene), 5 μl 10× PfuTurbo reaction buffer, 25 μM dNTPs, 0.5 μM of Illumina PE PCR primers (PE1.0, PE2.0). PCR cycling parameters were: 95 °C for 2 min, 98 °C for 30 s, then 10 cycles of 98 °C for 15 s, 65 °C for 30 s and 72 °C for 3 min, ending with one 72 °C for 5 min step. The reaction products were purified twice using Ampure-XT beads. Quality of the libraries and size distribution were assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies). Libraries were sequenced on Illumina MiSeq and HiSeq 2000 machines.

Targeted bisulphite library preparation and sequencing. Target regions were selected based on DNMT3A2 and DNMT3B1 enrichments. Primers were designed against the *in silico* converted templates using Primer3 (http://primer3.sourceforge.net) with slight modifications and subsequent selection of 96 bisulphite primer-pairs (product size: 200–400 bp) (Supplementary Table 1). Primers were commercially synthesized on 96-well plate format. A 2 μg sample of RNaseA-treated genomic DNA was mixed with 3.2 pM *Sss*I (NEB) methylated T7 and unmethylated lambda DNA and converted following standard Epitect bisulphite conversion kit (Qiagen) protocols. Bisulphite-converted DNA was amplified with specific primers with following cycling conditions: twenty touch-down cycles from 55 to 50 °C with 30 s at 95 °C, 30 s at 55/50 °C and 30 s at 72 °C, followed by forty cycles of 30 s at 95 °C, 30 s at 50 °C and 30 s at 72 °C. Then samples of 5 μl per reaction were pooled and purified using 2% agarose gels. Indexed libraries were prepared using the NEB next multiplex library kit (E73352L) according to manufacturer's protocol and sequenced on Illumina MiSeq machines with 150 bp paired-end reads.

Global quantification of 5-methylcytosine. Measurements of 5-methylcytosine levels were performed by Zymo Research (http://www.zymoresearch.com) using high-pressure liquid chromatography coupled to mass spectrometry.

Histone H3 isolation and mass spectrometry. Histones were extracted from ES cells using standard acidic extraction. H3 was isolated on 12% Tris-acetate gels (Novex). Amino groups were modified with propionic anhydride prior trypsin digestion. Differently methylated peptides were quantified by LC–MS on LTQ Orbitrap Velo, and peak area ratios were normalized to wild-type and histone H3.

Superimposition of PWWP structures. Modelling of BRPF1-PWWP with K36me3 (2X4W)[34], ZMYND11-PWWP with K36me3 (4N4H)[35] and DNMT3B-PWWP with Bis-Tris (1KHC)[36] was performed using PyMOL. We excluded the Bis-Tris molecule from 1KHC for better visualization.

ChIP-seq and bisulphite-seq reads processing. Individual ChIP-seq samples were filtered for low-quality reads and adaptor sequences, and mapped to the mouse genome (mm9) using the BOWTIE algorithm[51] allowing for two mismatches and mapping up to 100 different genomic sites per read. Reads were converted into the genomic ranges[37] format for downstream analysis in R. Reads that mapped to more than one genomic site were discarded, and remaining reads were shifted to account for sonication fragment length. For whole-genome bisulphite sequencing analysis, 100 bp and 150 bp reads were partitioned into 50 bp and pooled for subsequent alignment. Bisulphite alignments were performed using QuasR (http://www.bioconductor.org/packages/release/bioc/html/QuasR.html) in R with standard parameters for single-read bisulphite alignments. Methylation calls and read coverage per CpG were extracted and only CpGs covered more than ten times were used for subsequent analysis. CpGs overlapping with SNPs were removed. This resulted in 7.94, 7.35 and 9.29×10^6 CpG for *Dnmt*-TKO-DNMT3A2, *Dnmt*-TKO-DNMT3B1 and *Dnmt*-TKO-*Setd2*-KO-DNMT3B1, respectively. For targeted bisulphite sequencing, paired-end 150 bp reads were aligned using AmpliconBisSeq (https://github.com/

BIMSBbioinfo/AmpliconBiSeq) based on QuasR bisulphite alignment pipeline for paired-end reads. Methylation calls and coverage per CpG were extracted for the targeted regions using QuasR. Bisulphite conversion quality was confirmed by spiked-in controls of methylated T7 DNA and unmethylated lambda DNA.

Genomic coordinates. All genomic annotations are based on *Mus musculus* version NCBI37/mm9. Fully methylated regions (FMR), low methylated regions (LMR, $n = 8{,}249$) and unmethylated regions (UMR, $n = 3{,}117$) were obtained from a previous study[8]. Dense methylated regions (meth, $n = 3{,}626$) are based on CpG islands detected using the CpG cluster algorithm[38] and exceeding 80% average CpG methylation and min 350 bp size. RefSeq gene, exon and intron annotations were obtained from UCSC (http://genome.ucsc.edu). Repetitive regions were obtained using RepeatMasker annotations (UCSC) excluding low complexity repeats. Promoters were defined as 1 kb regions (−900 bp and +100 bp) around RefSeq TSS. For enrichments at genomic fractions, all overlaps between genomic regions were removed using the following hierarchy: promoters, exons, repeats and introns. The entire population of genomic instances defined above was used in the analysis. The large population sizes are adequate to measure effect size.

Genome-wide ChIP enrichments, identification of DNMT enriched regions and average profile calculations. In order to identify genome-wide enrichments for ChIP samples, the mouse genome was partitioned into 1-kb sized windows with an overlap of 500 bp. Windows overlapping with satellite repeats were removed in order to reduce false positives arising from differences in repeat number between mouse strains. Furthermore, windows with insufficient coverage in whole-genome bisulphite sequencing (less than 50% of all CpGs covered at least 10 times) were removed resulting in 1,995,777 windows. Enrichments were calculated as $\log_2$-fold changes over input chromatin after library size normalization and using a constant of eight pseudo-counts to reduce sampling noise. These genome-wide enrichments were subsequently used for clustering analysis using cross-correlation and for ranking genomic windows. In order to identify DNMT3A2 or DNMT3B1 enriched genomic regions, we first removed all genomic 1-kb windows with enrichment less than 0.5 in both replicates. The remainder was further used to calculate differentially enriched regions using DEseq[39]. Identified regions were used to calculate distribution of genomic and epigenomic features at DNMT bound sites and calculate DNA methylation in TKO cells expressing *de novo* DNMTs. Average profiles around the centre of non-overlapping genomic features (LMR, UMR, meth, TSS, CTCF, DNMT3-bound) are displayed as the mean of library size-normalized ChIP-seq signal and as average methylation of CpG for bisulphite data.

Expression-based analysis at gene bodies, exons and introns. For calculating enrichments at gene bodies, we used RefSeq annotations. First we removed all overlapping genes resulting in 13,273 unique gene regions. From this we further removed all genes shorter than 4 kb ($n = 3{,}016$). This excluded short open reading frames (ORFs) that contain large parts of promoter regions and CpG islands within their gene bodies that would preclude downstream analysis. To further exclude overlaps with promoter regions we defined all remaining gene bodies ($n = 10{,}257$) as regions 2 kb downstream of TSS to the end of the gene. For analysing exons and introns, only instances exceeding 80% CpG methylation were used to avoid overlaps with active proximal and distal regulatory sites. In addition, we split introns overlapping with exons and discarded instances shorter than 200 bp or with low coverage in whole-genome bisulphite experiments (less than 50% of all CpGs covered at least 10 times). ChIP enrichments were calculated as described above after normalization for feature length. For assessing transcriptional activity of exons, introns and gene bodies we mapped mRNA-seq reads data obtained from the same ES cell clone. Gene bodies, exons and introns were ranked by expression into five equally sized bins ($n = 2{,}051$ genes per bin). These bins were further used to calculate DNMT enrichments and average profiles. Correlations were calculated for all bins combined. Heatmap profiles around genes and k-means clustering based on the obtained profiles were generated using Genomation[40] in R. For comparison between ES cells and NP cells only gene bodies bound by DNMT3B in both or in either of the compared states were analysed ($n = 1{,}944$).

DNA methylation analysis. DNA methylation percentage, CpG and m-CpG densities per analysed regions were calculated as described previously[6,8]. Densities were calculated as the sum of SNP filtered CpGs per analysed region and normalized to 100 bp. Global re-methylation was calculated as percent of methylated cytosines within CpG sequence context for the entire library. For targeted bisulphite sequencing, only fragments and CpGs covered more than ten times were considered. Targets were chosen based on DNMT3A2, DNMT3B1 binding, H3K36me3 levels and DNA methylation percentage (Supplementary Table 1).

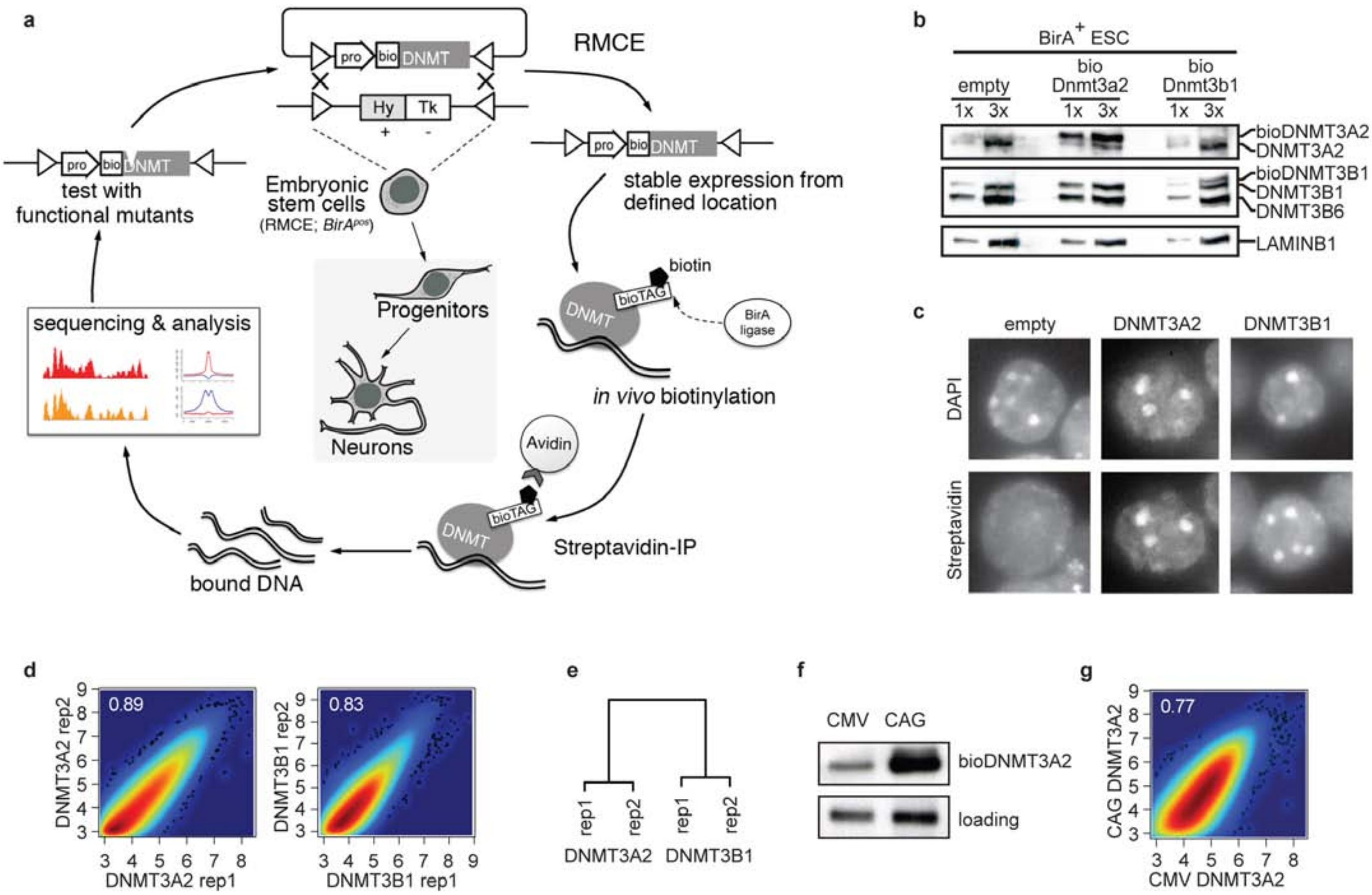
Distance–dependence of DNA methylation. Distance-dependent methylation analysis were performed only on CpGs covered at least ten times and more than 80% methylated in wild-type ES cells to exclude overlaps with active regulatory regions. For each of these CpGs, we first defined the distance to the nearest CpG and calculated the average methylation based on CpG-to-CpG distance. We further repeated this analysis using randomized methylation scores to test the obtained distance-dependence results. We repeated the distance-dependent methylation analysis on single DNA molecules using all reads mapping to chromosome 1. The methylation status of each pair of neighbouring CpGs on the same read was recorded. For each given distance between CpG pairs, the fraction of neighbouring methylated CpGs was then defined as the number of CpG pairs that are both methylated, divided by the number of CpGs where the upstream CpG is methylated and the downstream CpG is either methylated or unmethylated. For this analysis, only the first read of each paired-end sample was used. Distance-dependent correlation of methylation scores was performed for all CpGs methylated more than 80% in wild-type ES cells and covered at least ten times. Pairs of CpGs within the same distance were collected from the entire data set and Pearson's correlation coefficient was calculated for distances ranging from 2 to 400 bp.

Enrichments and DNA methylation around CTCF bound sites. CTCF sites were chosen based on similar binding to strong CTCF motifs in wild-type ES and TKO cells[8]. Only CTCF sites outside of promoter regions were used ($n = 4{,}126$). MNaseI data from ES cells[41] was used to display nucleosome occupancy around these sites in wild-type cells. For correct display of CTCF ChIP-seq location, reads were shifted based on their forward/reverse orientation.

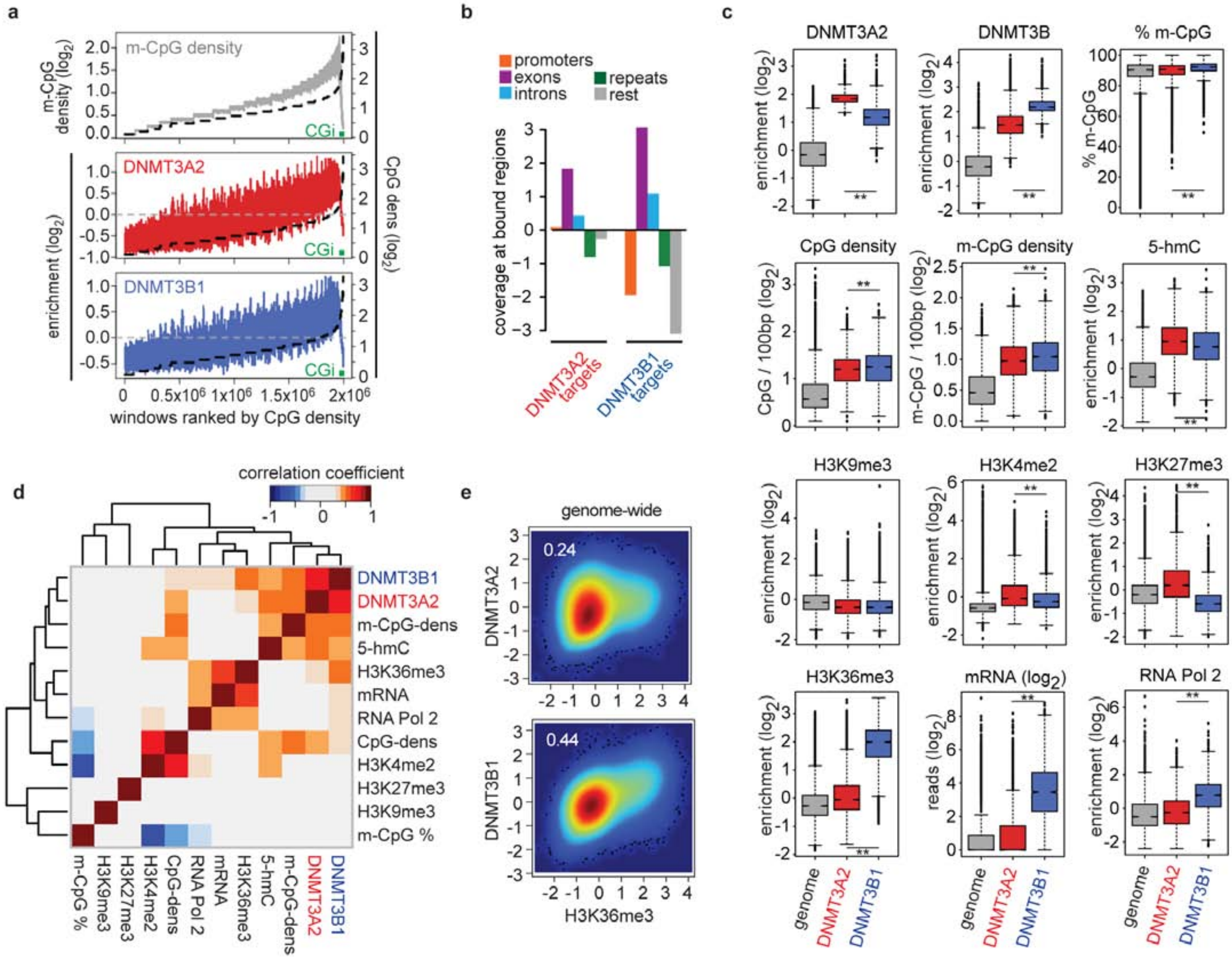
Prediction of DNMT binding based on linear models. In order to test binding prediction of DNMTs based on genomic features, we have chosen a set of 1-kb sized windows equally representing promoters, gene bodies, and intergenic regions ($n = 60{,}000$). Individual models were build using the *lm()* function in R.

Data sets used in this study. Data sets generated for this study, DNMT ChIP-seq and wgBis-seq are deposited at GEO under the accession number GSE57413. Data for ES cell and NP cell H3K36me3, Pol II are available from[14], Input, wild-type ES cell wgBis-seq, H3K4me2, H3K27me3, mRNA and CTCF binding in wild-type and TKO cells[8], ES cell 5-hmC-GLIB[23], ES cell serum H3K4me3 and H3K9me3 (ref. 42), MNaseI[41].

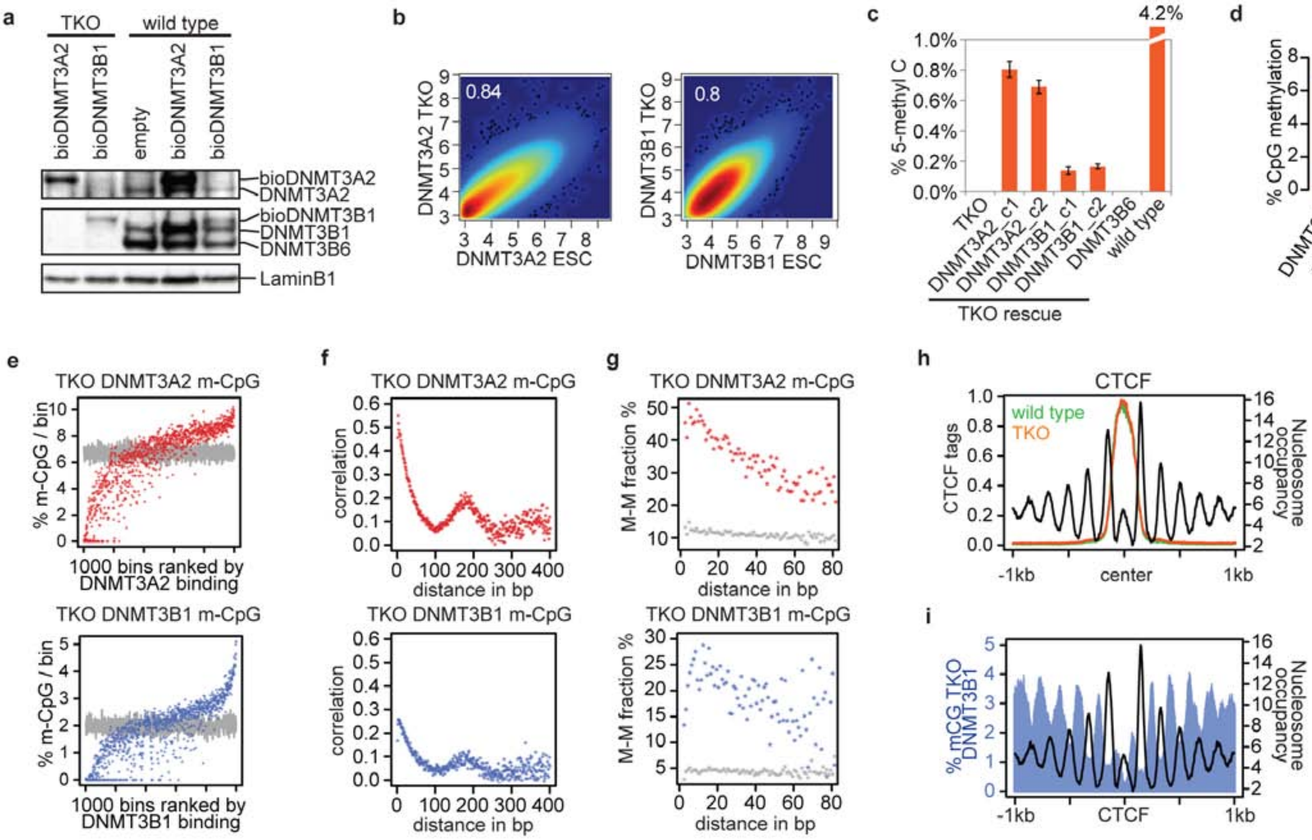
31. Bibel, M., Richter, J., Lacroix, E. & Barde, Y.-A. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nature Protocols* **2,** 1034–1043 (2007).
32. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
33. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnol.* **31,** 827–832 (2013).
34. Vezzoli, A. *et al.* Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nature Struct. Mol. Biol.* **17,** 617–619 (2010).
35. Wen, H. *et al.* ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature* **508,** 263–268 (2014).
36. Wu, H. *et al.* Structural and histone binding ability characterizations of human PWWP domains. *PLoS ONE* **6,** e18919 (2011).
37. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9,** e1003118 (2013).
38. Hackenberg, M. *et al.* CpGcluster: a distance-based algorithm for CpG-island detection. *BMC Bioinformatics* **7,** 446 (2006).
39. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).
40. Akalin, A., Franke, V., Vlahoviček, K., Mason, C. E. & Schübeler, D. genomation: a toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics*, (2014).
41. Teif, V. B. *et al.* Genome-wide nucleosome positioning during embryonic stem cell development. *Nature Struct. Mol. Biol.* **19,** 1185–1192 (2012).
42. Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149,** 590–604 (2012).
43. Chen, T., Tsujimoto, N. & Li, E. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. *Mol. Cell. Biol.* **24,** 9048–9058 (2004).
44. Bachman, K. E., Rountree, M. R. & Baylin, S. B. Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J. Biol. Chem.* **276,** 32282–32287 (2001).
45. Untergasser, A. *et al.* Primer3–new capabilities and interfaces. *Nucleic Acids Res.* **40,** e115 (2012).
46. Jung, H. R., Pasini, D., Helin, K. & Jensen, O. N. Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. *Mol. Cell. Proteomics* **9,** 838–850 (2010).
47. Ferrari, K. J. *et al.* Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol. Cell* **53,** 49–62 (2014).
48. Zhang, Y. *et al.* H3K36 histone methyltransferase Setd2 is required for murine embryonic stem cell differentiation toward endoderm. *Cell Reports* **8,** 1989–2002 (2014).
49. Pei, J., Kim, B. H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36,** 2295–2300 (2008).
50. Jeong, S. *et al.* Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing Methylated DNA. *Mol. Cell. Biol.* **29,** 5366–5376 (2009).
51. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).



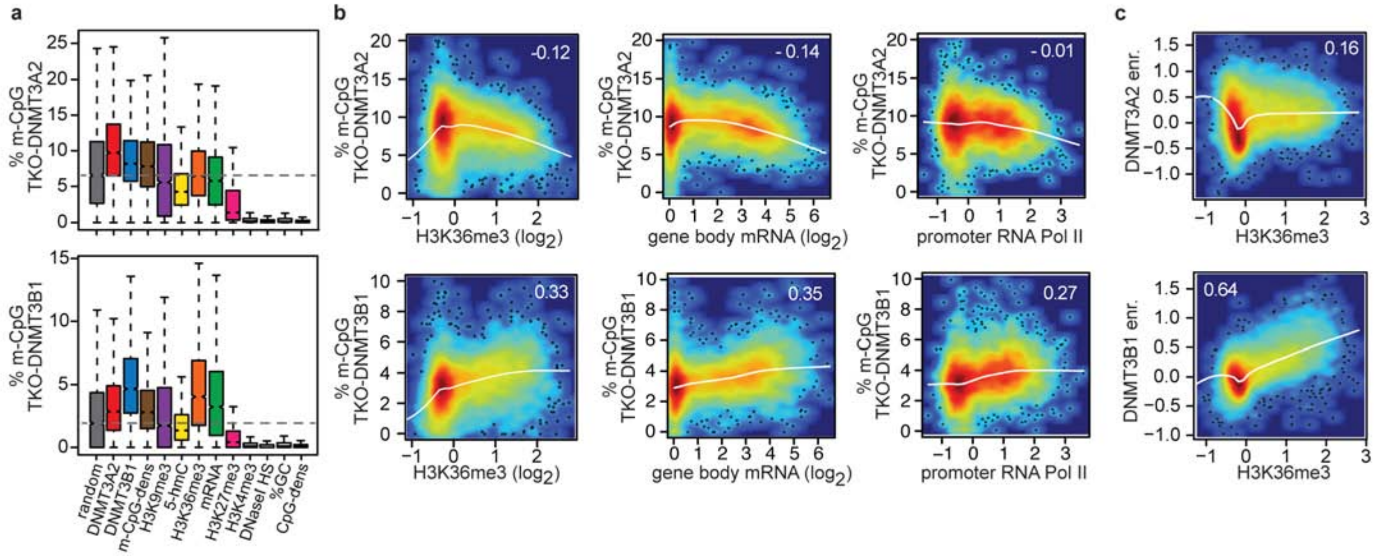
Extended Data Figure 1 | Biotin tagging of DNMT3A2 and DNMT3B in mouse embryonic stem cells. a, Outline of the recombinase-assisted mapping of biotin-tagged proteins (RAMBiO) approach[6]. Expression constructs are inserted at a defined genomic site using recombinase-mediated cassette exchange (RMCE). Tagged proteins are biotinylated *in vivo* by the constitutively expressed biotin ligase Bir-A. ChIP is performed using streptavidin-based immunoprecipitation. Protein variants, inserted at the same genomic locus and under the same promoter can be compared to their wild-type counterparts in a controlled manner. **b**, Immunoblot for DNMT3A2 and DNMT3B1 to monitor expression levels of endogenous and biotinylated DNMT proteins. Detection was performed with antibodies against DNMT3A and DNMT3B using whole ES cell extracts loaded in two different concentrations (1× and 3×). LaminB1 served as a loading control. **c**, Nuclear localization of biotin-tagged DNMT3A2 and DNMT3B1 detects localization at DAPI-dense chromocentres, as previously reported for their endogenous counterparts[43,44]. Detection of biotinylated proteins was performed by streptavidin-AF568. **d**, Scatterplots showing genome-wide correlations between replicates of DNMT3A2 and DNMT3B1 calculated based on $\log_2$-transformed read counts in 1-kb sized windows covering the entire genome but excluding highly repetitive regions (n = 1,995,777 windows). Pearson's correlation is shown. **e**, Dendrogram shows hierarchical clustering of genome-wide DNMT3A2 and DNMT3B1 enrichments over input in 1-kb sized windows covering the entire genome. **f**, Immunoblot showing protein levels of biotin-tagged DNMT3A2, expressed from the RMCE site using CAG or CMV promoters. **g**, Scatterplot showing correlation in genome-wide binding between DNMT3A2 expressed at different levels (CAG and CMV) calculated based on $\log_2$-transformed read counts in 1-kb sized windows covering the entire genome (n = 1,995,777 windows). This indicates similar binding, despite different expression levels.



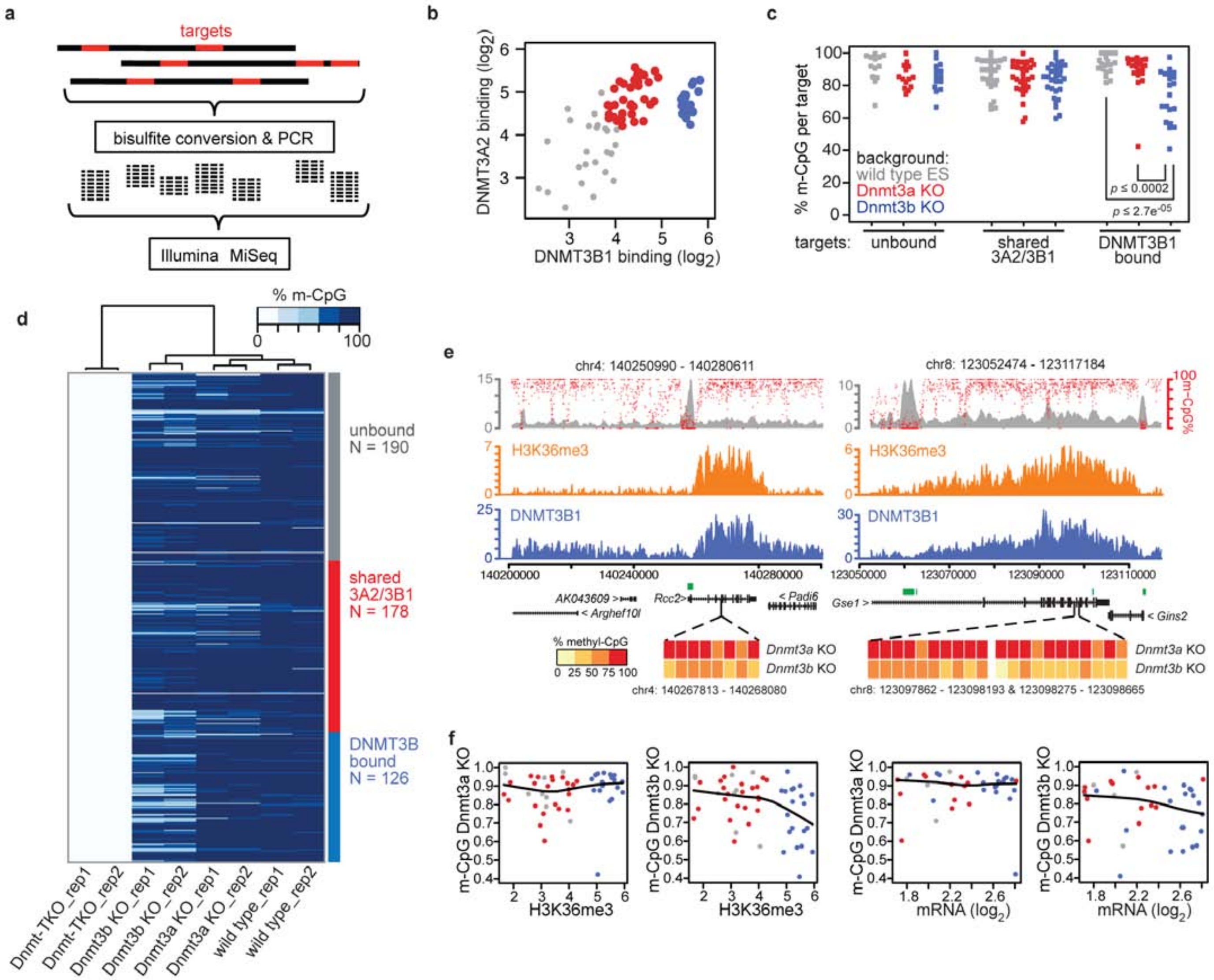
Extended Data Figure 2 | Genomic binding preferences of biotin-tagged DNMTs. **a**, DNMT3A2 and DNMT3B1 enrichment follows methyl-CpG density. Shown are methyl-CpG density per 100 bp (grey) and enrichments for DNMT3A2 (red) or DNMT3B1 (blue) at 1-kb windows of the entire genome ranked by CpG density (n = 1,995,777 windows). Dashed line displays $\log_2$-transformed CpG density. Smoothing was applied using 200 consecutive windows. Genomic windows overlapping with unmethylated CpG islands are indicated by a green square. **b**, Enrichment of genomic features at DNMT3A2 and DNMT3B1-bound sites indicate an increased preference for gene bodies by DNMT3B. Shown are enrichments of different genomic features within DNMT-enriched sites (DNMT3A2, n = 6,499; DNMT3B1, n = 4,799) compared to their average genomic distribution. **c**, Boxplots showing the distribution of various genomic and epigenomic features at sites preferentially bound by DNMT3A2 (red, n = 6,499) or DNMT3B1 (blue, n = 4,799). Genome-wide distribution is shown in grey for comparison. Box indicates the interquartile range (IQR) and whiskers denote the 1.5× IQR range. Statistical differences between DNMT3A2 and DNMT3B1 are calculated using Mann–Whitney–Wilcoxon (MWW) tests (**$P < 2.2 \times 10^{-16}$). Strongest differences between DNMT3A2 and DNMT3B1-enriched sites are found for H3K36me3 and mRNA. **d**, Unsupervised clustering highlights similarities and individual preferences of the *de novo* DNMTs to genomic sites with unique chromatin and DNA methylation features extracted from publicly available data sets. Cross-correlation matrix is calculated based on Pearson's correlation coefficient derived from 1-kb windows covering chromosome 19 (n = 100,477). **e**, Genome-wide correlations between DNMT3A2 or DNMT3B1 enrichment over input ($\log_2$) and H3K36me3 enrichment over input ($\log_2$) using 1-kb windows covering the entire genome (n = 1,995,777). Pearson's correlation is shown.



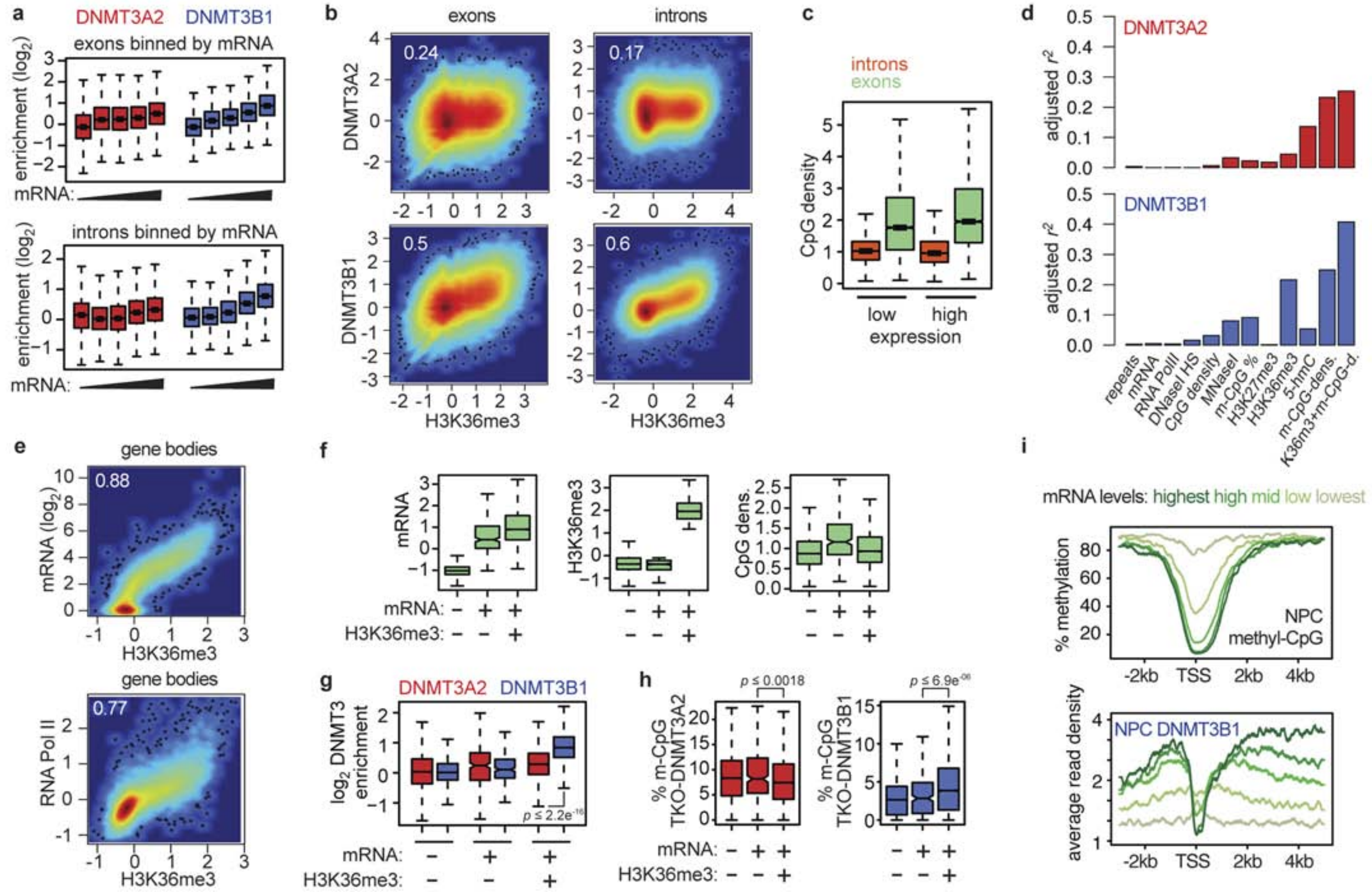
Extended Data Figure 3 | *De novo* methylation follows binding preferences of DNMTs and indicates processivity and substrate accessibility. **a**, Immunoblot showing expression levels of biotinylated DNMT3A2 and DNMT3B1 proteins in wild-type and TKO cells. Detection was performed using antibodies against DNMT3A or DNMT3B on whole cell extracts. Lamin B1 serves as loading control. Empty indicates wild-type Bir-A cells without biotin-tagged DNMTs. **b**, Scatterplots showing correlations between wild-type ES cells and TKO ES cells for DNMT3A2 and DNMT3B1 binding, respectively. Pearson's correlation coefficients are calculated based on $\log_2$-transformed read counts in 1-kb windows covering the genome ($n = 1{,}995{,}777$). **c**, Global 5-methylcytosine levels measured by HPLC-MS in *Dnmt*-TKO cells, TKO cells with reintroduced DNMT3 isoforms and wild-type ES cells. Displayed are the fractions of methylated cytosines from two biological replicates for DNMT3A2 and DNMT3B1, respectively. DNMT3B6, a catalytic inactive isoform served as a negative control. Error bars indicate standard deviation calculated from three technical replicate measurements. **d**, Percentage of methylated cytosines (CpG only) in TKO cells expressing DNMT3A2 (red) or DNMT3B1 (blue) respectively. Calculation is based on whole-genome bisulphite sequencing using CpGs covered at least ten times (DNMT3A2, $n = 7.94 \times 10^6$ CpGs; DNMT3B1, $n = 7.35 \times 10^6$ CpGs). Dashed line indicates non-conversion rates based on spiked-in, unmethylated lambda phage DNA. **e**, *De novo* methylation follows genomic binding of DNMT3A2 and DNMT3B1. Shown is median DNA methylation calculated using 1-kb sized windows ranked by DNMT3A2 or DNMT3B1 enrichments. Equal numbers of genomic windows were binned (433 per bin) after DNMT3-enrichment ranking, and median methylation was calculated per bin ($n = 1{,}000$ bins). Only windows where at least 50% CpGs were covered more than 10 times in both experiments were used. **f**, Distance-dependent correlation of methylation levels between neighbouring CpGs indicates similar methylation for pairs of CpGs in close proximity and those at a ~180 bp distance. **g**, Single-molecule analysis shows distance-dependent methylation of neighbouring CpGs on the same DNA strand. Shown is the fraction of neighbouring CpGs at a given distance, where both consecutive CpGs on the same DNA molecule are methylated. Grey dots indicate the same analysis but using randomized methylation states between reads as a control. **h**, Average density plots centred around strong CTCF binding sites ($n = 4{,}126$ sites) showing CTCF binding in wild-type (green) and TKO cells (orange)[8]. Nucleosome occupancy surrounding these CTCF sites is shown in black and was calculated based on MNaseI-seq in mouse ES cells[41]. **i**, *De novo* methylation in TKO cells occurs preferentially at linker regions between positioned nucleosomes, here illustrated around genomic sites bound by CTCF. Shown is average DNA methylation by DNMT3B1 in TKO cells.



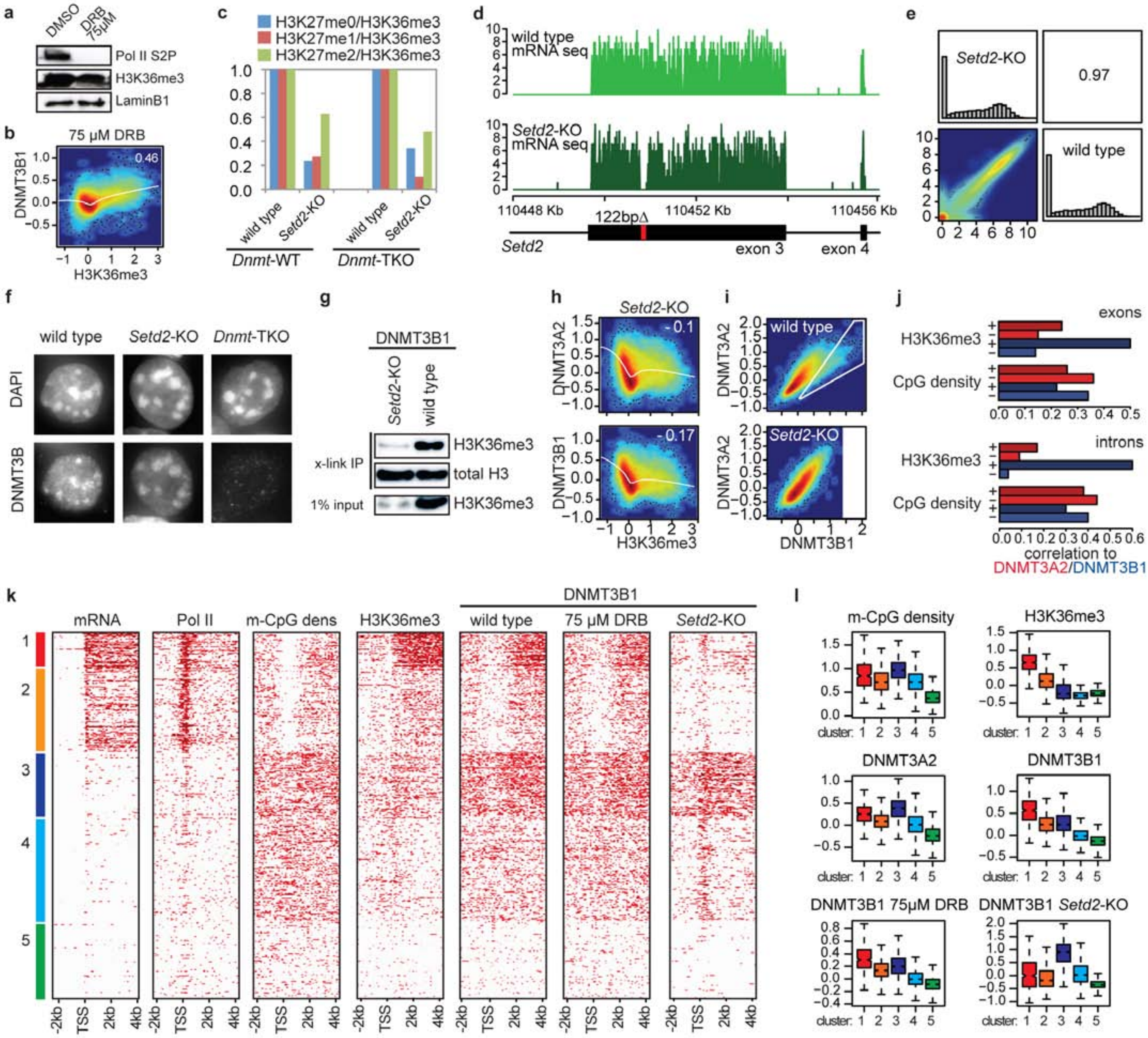
Extended Data Figure 4 | *De novo* DNA methylation and binding of DNMT3B to gene bodies correlates with transcriptional activity. **a**, Boxplots showing the distribution of *de novo* DNA methylation at genomic regions bound by DNMT3A2 ($n = 6{,}499$), DNMT3B1 ($n = 4{,}799$) and genomic windows ($n = 4{,}000$, size = 1 kb) enriched for various chromatin and sequence-based features. Regions were selected as top 1% genomic windows enriched for the indicated feature and sufficiently covered in both bisulphite sequencing experiments. Grey boxplot indicates *de novo* methylation using a set of randomized genomic windows covered in both bisulphite sequencing experiments ($n = 4{,}000$ windows, size = 1 kb). The median of this randomized sample sets the baseline (dashed grey line). **b**, Relationship between transcriptional activity and *de novo* DNA methylation at gene bodies. Scatterplots comparing *de novo* DNA methylation in TKO cells expressing DNMT3A2 or DNMT3B1 with H3K36me3, RNA pol II enrichment or mRNA levels at gene bodies. Only gene bodies, sufficiently covered in both bisulphite sequencing experiments were used ($n = 5{,}686$). **c**, Scatterplot indicating binding of DNMT3B to actively transcribed gene bodies ($n = 10{,}257$). Shown is DNMT3 enrichment compared to H3K36me3 enrichment at gene bodies in $\log_2$ scale. Pearson's correlation coefficient and Loess curves are shown (**b** and **c**).



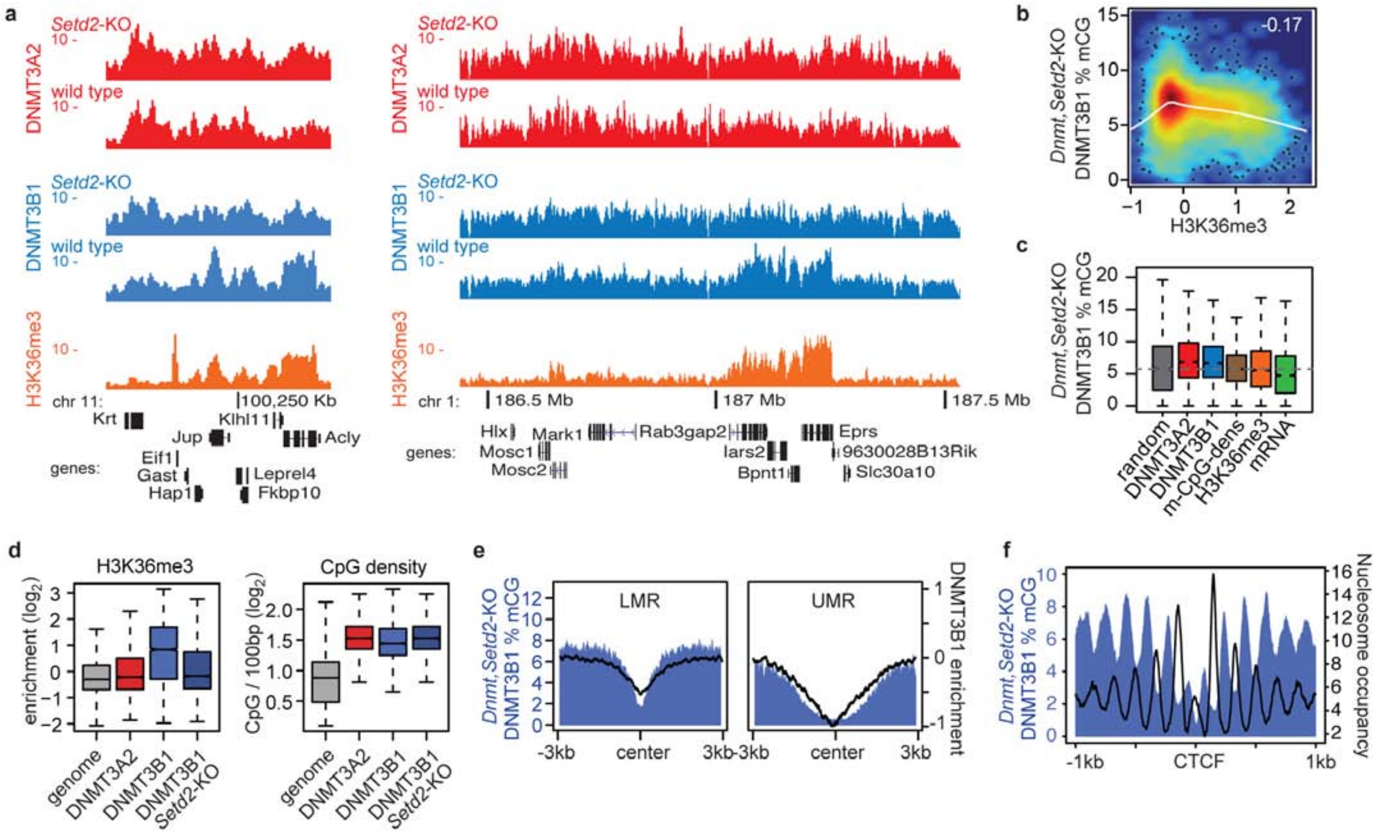
Extended Data Figure 5 | Methylation analysis at selected sites indicates a role for DNMT3B in methylation fidelity at transcribed genes. **a**, Scheme illustrating the targeted bisulphite sequencing approach. Primers for selected genomic regions (**b**) were designed using Primer3 (ref. 45), genomic DNA from wild-type and mutant ES cells was PCR-amplified after bisulphite conversion. PCR products were pooled and sequenced on an Illumina MiSeq machine. **b**, Selected genomic regions analysed by targeted bisulphite sequencing (Supplementary Table 1). The following sites were selected: methylated regions with reduced binding of DNMT3A2 and DNMT3B1 (unbound, $n = 25$, grey), shared binding by DNMT3A2 and DNMT3B1 (shared 3A2/3B1, $n = 35$, red), and preferential binding of DNMT3B1 (DNMT3B bound, $n = 21$, blue). Shown are $\log_2$-transformed reads indicating DNMT3A2/3B1 binding at the selected regions. **c**, DNA methylation at a subset of transcriptionally active, DNMT3B1-bound sites is reduced in *Dnmt3b*-KO, but not in *Dnmt3a*-KO ES cells. Shown are average methylation values from bisulphite PCR analysis at the genomic sites defined in **b** and sufficiently covered in bisulphite sequencing experiments. Analysis was performed in wild-type ES cells (grey) and ES cells lacking DNMT3A (red) or DNMT3B (blue). *P* values were calculated using MWW tests. **d**, Heatmap showing DNA methylation at single CpGs within analysed regions for different mutants including *Dnmt*-TKO cells as control (two technical replicates each). Clustering was performed based on methylation of single CpGs. **e**, Genomic examples for sites with reduced DNA methylation in *Dnmt3b*-KO cells. Top track shows endogenous methylation in wild-type ES cells (red dots) and CpG density per 100 bp (grey). H3K36me3 and DNMT3B1 binding are shown as library-normalized read counts per 100 bp. Percent DNA methylation per individual CpGs within amplified regions in *Dnmt3a*-KO and *Dnmt3b*-KO cells is shown below. **f**, Average DNA methylation at analysed sites in *Dnmt3a* or *Dnmt3b* KO cells compared to H3K36me3 enrichment or mRNA levels ($\log_2$-transformed) in wild-type ES cells. Black line indicates average as Loess curves.



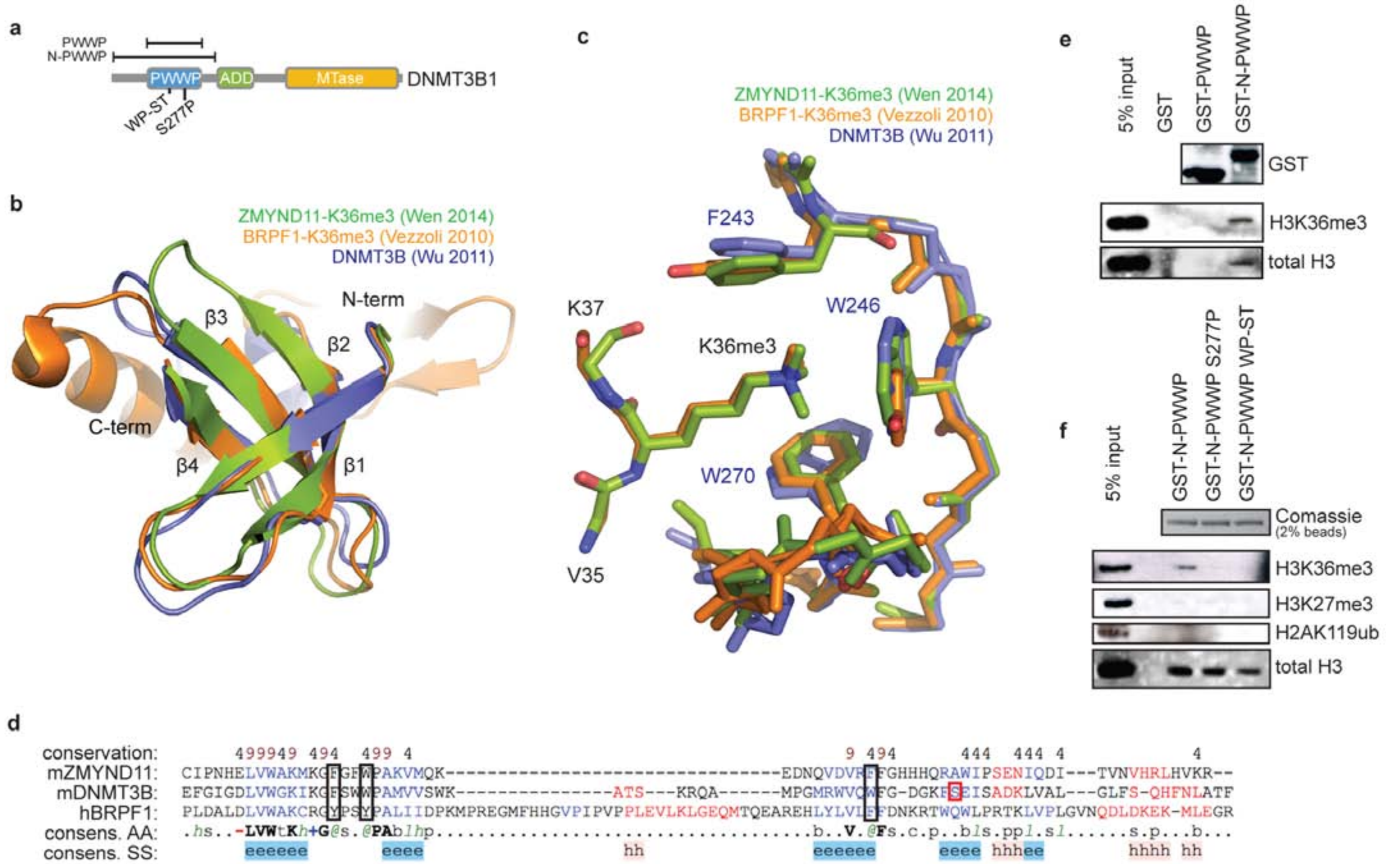
Extended Data Figure 6 | H3K36me3 guides binding of DNMT3B to active gene bodies. **a**, Boxplots displaying enrichment for DNMT3A2 and DNMT3B1 at exons and introns that are grouped according to mRNA levels. Boxes denote the IQR and the lowest/highest values within 1.5× IQR are shown as whiskers. **b**, Scatterplots comparing H3K36me3 enrichment ($\log_2$) to DNMT3A2 and DNMT3B1 enrichments ($\log_2$) separately for exons and introns. Pearson's correlation coefficient is shown. **c**, Boxplots showing distribution of CpG densities (calculated per 100 bp) for exons and introns with lowest and highest expression. Shown are the first and fifth quantiles from **a**, based on mRNA levels. **d**, Adjusted r^2 indicates prediction performance for DNMT3A2 and DNMT3B1 binding based on various genomic and epigenomic features at selected genomic sites. Prediction was performed using linear models on individual and combined values over a defined set of genomic regions (60,000 genomic sites of 1 kb, see Methods). **e**, Scatterplot comparing H3K36me3 enrichment at gene bodies ($n = 10{,}257$) with RNA pol II enrichment or mRNA levels ($\log_2$). Pearson's correlation coefficient is indicated. **f**, Boxplots showing distribution of mRNA levels, H3K36me3 enrichment over input and CpG density per 100 bp at introns that are not transcribed (mRNA (−) and H3K36me3 (−), $n = 41{,}656$), transcribed but are not H3K36me3 positive (mRNA (+) and H3K36me3 (−), $n = 267$) or transcribed and H3K36me3 positive (mRNA (+) and H3K36me3 (+), $n = 11{,}670$). Boxes denote the IQR and lowest/highest values within 1.5× IQR are shown as whiskers. **g**, Boxplots for same introns fractions as in **f**, but showing DNMT3A2 and DNMT3B1 enrichments over input ($\log_2$). P values were calculated using MWW tests. **h**, Boxplots indicating *de novo* DNA methylation in *Dnmt*-TKO cells expressing either DNMT3A2 (red) or DNMT3B1 (blue) for the same intron fractions as in **f** and **g**. **i**, Average profiles around TSS, grouped based on transcriptional activity in neuronal progenitor cells (NP cells, 2,051 genes per bin). Shown is average DNA methylation in wild-type NP cells and average read counts indicating bio-DNMT3B1 binding in NP cells.



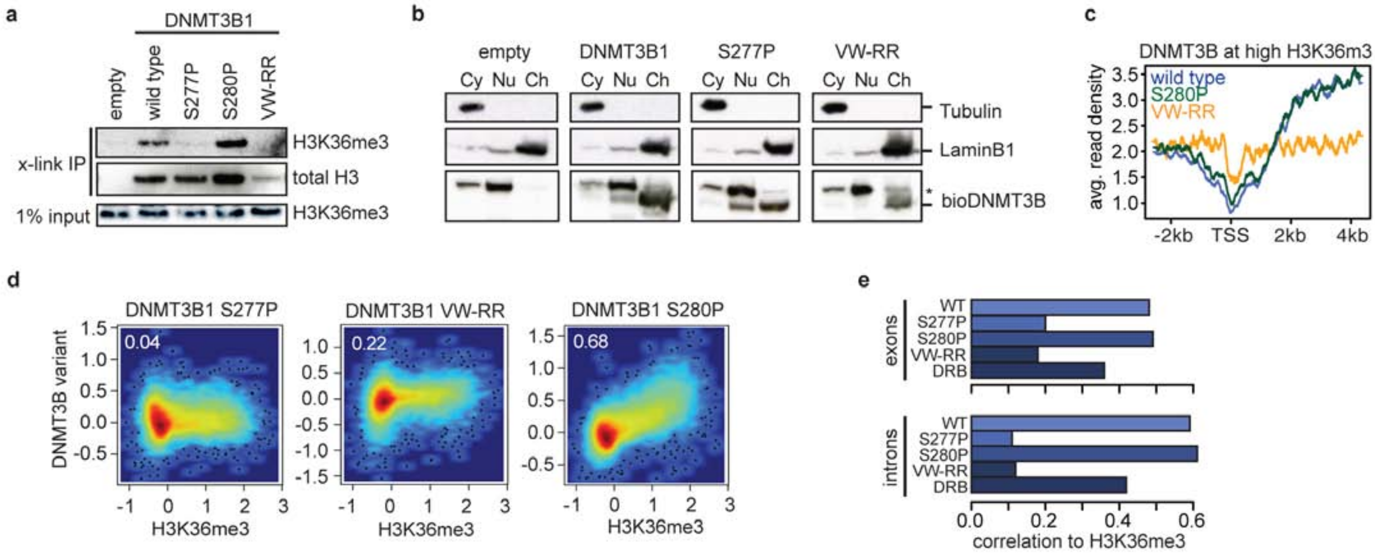
Extended Data Figure 7 | Binding of DNMT3B to gene bodies is dependent on SETD2. **a**, Immunoblot illustrating absence of RNA pol II phosphorylation at serine 2 (S2P) upon treatment with 75 μM DRB for 10 h. H3K36me3 is still present after this block in elongation. **b**, DNMT3B1 binds active genes in the absence of acute transcription. Scatterplots comparing H3K36me3 and DNMT3B1 enrichments ($\log_2$) at gene bodies in ES cells after 10 h treatment with 75 μM DRB. **c**, Relative quantification by MS shows reduction of H3K36me3 in *Setd2*-KO ES cells. Data are shown relative to the corresponding wild-type background (*Dnmt*-wild type or *Dnmt*-TKO cells). Measurements for H3K36me3 in combination with H3K27 methylation on the same tryptic peptide are shown. Previous quantitative MS[46] and genome-wide measurements[47] indicate that H3K36me3/K27me0 and H3K36me3/K27me1 overlap in mouse ES cells. **d**, CRISPR/Cas9 mediated deletion in exon 3 of the *Setd2* gene, identified by mRNA-seq in wild type and *Setd2*-KO cells. **e**, mRNA levels of highly expressed genes are not severely altered in absence of SETD2, as previously observed[48]. Scatterplot shows $\log_2$-transformed mRNA read counts mapping to transcripts in wild type and *Setd2*-KO cells. Histograms indicate distribution of mRNA levels. Pearson's correlation coefficient is shown. **f**, Localization of endogenous DNMT3B to centromeric regions is not altered in absence of SETD2. *Dnmt*-TKO is shown as control. Detection was performed using antibodies specific to DNMT3B. **g**, Chromatin immunoprecipitation of biotinylated DNMT3B1 followed by histone modification analysis by immunoblot. H3K36me3-chromatin is precipitated only in *Setd2* wild-type cells, whereas lack of SETD2 does not alter global chromatin precipitation (measured by total H3 recovery). **h**, Deletion of *Setd2* leads to loss of preferential binding of DNMT3B1 to active genes, but does not alter DNMT3A2 binding (see Extended Data Fig. 4c). Pearson's correlation coefficients and trend lines calculated using Loess fit are shown. **i**, Scatterplots comparing DNMT3B with DNMT3A2 binding to gene bodies in wild-type and *Setd2*-KO cells. This indicates similarities in binding in absence of H3K36me3. Highlighted gene bodies indicate H3K36me3-dependent binding of DNMT3B in wild-type cells. DNMT3B enrichment in wild-type and *Setd2*-KO cells are shown at similar scale for better comparison. **j**, DNMT3B1 binding at H3K36me3-positive exons and introns depends on SETD2 (compare to Fig. 3a). Pearson's correlation coefficient between DNMT3A2 (red) or DNMT3B1 (blue) enrichments and H3K36me3 or CpG densities at exons and introns is calculated for wild-type (+) and for *Setd2*-KO (−) cells. Note that CpG density-dependent binding is not altered. **k**, Heatmaps of mRNA levels, Pol II occupancy, DNA methylation density, H3K36me3 and DNMT3B1 levels at all non-overlapping genes longer than 4 kb ($n = 10,257$). Genes were separated in five clusters based on k-means clustering. **l**, Boxplots indicate properties of gene bodies clustered based on **k**. DNMT3B1 binding to individual transcribed gene bodies (clusters 1 and 2) is detectable in wild-type and DRB-treated cells, but lost in *Setd2*-KO cells. Cluster 3 contains gene bodies with elevated CpG densities that are bound by DNMT3A2 and DNMT3B1 independent of H3K36me3.



Extended Data Figure 8 | Targeting of *de novo* methylation to H3K36me3 sites by DNMT3B is impaired in *Setd2*-KO cells. **a**, Genomic regions illustrating reduced binding of DNMT3B1 to active genes in absence of SETD2 and H3K36me3, whereas DNMT3A2 remains unchanged. Shown are read counts per 100 bp. Location of CpG islands (UCSC) and genes (RefSeq) are displayed below. **b**, Deletion of *Setd2* in *Dnmt*-TKO cells (*Dnmt, Setd2*-KO) that express DNMT3B1 results in loss of *de novo* methylation of sites that were positive for H3K36me3 (Compare to Extended Data Fig. 4b). Scatterplot comparing *de novo* DNA methylation in *Dnmt, Setd2*-KO cells expressing DNMT3B1 to H3K36me3 enrichments ($\log_2$) at gene bodies in wild-type cells. Pearson's correlation and trend line (Loess) are shown. **c**, Boxplots illustrating the distribution of *de novo* methylation by DNMT3B1 in *Dnmt*, *Setd2*-KO cells at different genomic regions (compare to Extended Data Fig. 4a). **d**, Boxplots indicating requirement for SETD2 in DNMT3B-dependent *de novo* methylation at H3K36me3 sites, but not at CpG-dense sites. Shown are H3K36me3 enrichments and CpG densities at sites preferentially methylated in *Dnmt*-TKO cells expressing DNMT3A2 or DNMT3B1 and *Dnmt, Setd2*-KO cells expressing DNMT3B1. Grey box denotes genome-wide distribution of H3K36me3 and CpG density for comparison. Boxes denote IQR and whiskers 1.5× IQR. **e**, Protection from *de novo* methylation of LMRs and UMRs is also observed in *Dnmt*, *Setd2*-KO cells expressing DNMT3B1. Shown are average density plots centred around methylation segments[8]. Black line indicates DNMT3B1 enrichments in *Dnmt*-TKO cells. **f**, Preferential *de novo* methylation at linker regions around sites bound by CTCF in DNMT3B1-expressing *Dnmt, Setd2*-KO cells. Shown are average density plots for DNA methylation in *Dnmt, Setd2*-KO cells (blue) and positioned nucleosomes in wild-type ES cells (black) centred at strong CTCF sites ($n = 4,126$).



Extended Data Figure 9 | The N-terminal part of DNMT3B interacts with H3K36me3-modified nucleosomes and requires a functional PWWP domain. **a**, Protein model for DNMT3B indicating known domains and mutations introduced in this study. **b**, Superimposition of published PWWP domain structures from DNMT3B with Bis-Tris in blue (Bis-Tris not shown)[36], ZMYND11 in presence of K36me3 in green[35] and BRPF1 in presence of K36me3 in orange[34] shows structural similarities between the analysed PWWP domains. **c**, Close-up of the aromatic cage involved in K36me3 interactions. Same structures as in **b**. **d**, Comparison of amino acid sequences of DNMT3B, BRPF1 and ZMYND11 PWWP domains using PROMALS3D[49] based on structure PDB IDs (1KHC, 2X4W and 4N4H). Conservation index is shown above (max score = 9), amino acid (AA) and secondary structure (SS) consensus is shown below. Conserved amino acids are indicated in bold, aromatic amino acids are indicated as the @ symbol. Consensus structures: *h* = α-helix, *e* = β-sheet. Residues of the aromatic cage involved in K36me3 recognition and one mutated in ICF syndrome patients (S277P) are highlighted by black and red rectangles, respectively. **e**, HeLa mononucleosome interaction assays using GST-fused PWWP domain alone (amino acids 227 to 343) and N-terminal part of DNMT3B containing the PWWP domain amino acids 1 to 427 (N-PWWP) (see **a**). Interactions with H3K36me3 nucleosomes are observed only for the extended PWWP domain. In line with this finding, deletion of the N-terminal part of DNMT3B, preceding the PWWP domain has previously been shown to disrupt anchoring to nucleosomes[50]. GST alone is used as control. **f**, The PWWP domain in the N-terminal DNMT3B construct is required for interactions with H3K36me3-modified nucleosomes. The ICF (S277P) mutation and substitutions of residues of the aromatic cage (WP-ST) disrupt interactions with H3K36me3 modified nucleosomes. Nucleosomes modified at H3K27me3 or H2AK119ub were not enriched in the pull-downs.



Extended Data Figure 10 | Mutations in the PWWP domain of DNMT3B1 disrupt chromosomal binding and preference for H3K36me3. **a**, Immunoblot of wild-type and PWWP mutant DNMT3B1 proteins after bio-ChIP indicates that a functional PWWP domain is required for interactions with chromatin modified at H3K36me3. Mutations introduced in the PWWP domain (S277P/ICF and substitutions in the conserved residues VW) abolish interactions with H3K36me3. **b**, Cellular fractionation into cytosolic (cy), soluble nuclear (nu) extracted at 100 mM NaCl and chromatin (ch) extracted at 500 mM NaCl indicates that introduced mutations do not severely affect nuclear localization or chromatin association of the biotin-tagged DNMT3B protein. Detection was performed using streptavidin. Asterisk indicates endogenous biotinylated proteins. Tubulin and LaminB1 served as loading controls. **c**, Substitutions in conserved PWWP domain residues (VW-RR, orange) result in loss of binding to active gene bodies (here shown at genes with high intragenic H3K36me3 levels, $n = 2{,}051$). A mutation within a flexible domain (S280P, green), three amino acids downstream of the ICF mutation, has no effect on H3K36me3 interactions, suggesting that the S277 residue, mutated in ICF patients is critical for this interaction (compare to Fig. 4d). **d**, Scatterplots showing the relation between H3K36me3 enrichments and enrichment of DNMT3B1 mutant proteins at gene bodies in wild-type ES cells. Pearson's correlation coefficient is shown. **e**, Pearson's correlation coefficient for DNMT3B1 mutant protein enrichments versus H3K36me3 levels at exons or introns (compare to Fig. 3a). 'DRB' indicates wild-type DNMT3B in cells treated with the pol II elongation inhibitor DRB (10 h, 75 μM).

doi:10.1038/nature14337

The Paf1 complex represses small-RNA-mediated epigenetic gene silencing

Katarzyna Maria Kowalik[1,2]*, Yukiko Shimada[1,2]*, Valentin Flury[1,2], Michael Beda Stadler[1,2,3], Julia Batki[4] & Marc Bühler[1,2]

RNA interference (RNAi) refers to the ability of exogenously introduced double-stranded RNA to silence expression of homologous sequences. Silencing is initiated when the enzyme Dicer processes the double-stranded RNA into small interfering RNAs (siRNAs). Small RNA molecules are incorporated into Argonaute-protein-containing effector complexes, which they guide to complementary targets to mediate different types of gene silencing, specifically post-transcriptional gene silencing and chromatin-dependent gene silencing[1]. Although endogenous small RNAs have crucial roles in chromatin-mediated processes across kingdoms, efforts to initiate chromatin modifications *in trans* by using siRNAs have been inherently difficult to achieve in all eukaryotic cells. Using fission yeast, here we show that RNAi-directed heterochromatin formation is negatively controlled by the highly conserved RNA polymerase-associated factor 1 complex (Paf1C). Temporary expression of a synthetic hairpin RNA in Paf1C mutants triggers stable heterochromatin formation at homologous loci, effectively silencing genes *in trans*. This repressed state is propagated across generations by the continual production of secondary siRNAs, independently of the synthetic hairpin RNA. Our data support a model in which Paf1C prevents targeting of nascent transcripts by the siRNA-containing RNA-induced transcriptional silencing complex and thereby epigenetic gene silencing, by promoting efficient transcription termination and rapid release of the RNA from the site of transcription. We show that although compromised transcription termination is sufficient to initiate the formation of bi-stable heterochromatin by *trans*-acting siRNAs, impairment of both transcription termination and nascent transcript release is imperative to confer stability to the repressed state. Our work uncovers a novel mechanism for small-RNA-mediated epigenome regulation and highlights fundamental roles for Paf1C and the RNAi machinery in building epigenetic memory.

In the fission yeast *Schizosaccharomyces pombe*, a functional RNAi pathway is required for the formation and stable propagation of constitutive heterochromatin found at pericentromeric repeat sequences. *S. pombe* contains single genes encoding for an Argonaute and a Dicer protein, called *ago1+* and *dcr1+*, respectively. Centromeres of *ago1Δ* or *dcr1Δ* cells have markedly reduced histone 3 lysine 9 (H3K9) methylation, which is a hallmark of heterochromatin, and defective chromosome segregation and heterochromatic gene silencing[2]. Ago1 is loaded with endogenous small RNAs corresponding to heterochromatic repeats, and interacts with Chp1 and Tas3 to form the RNA-induced transcriptional silencing (RITS) complex[3]. Current models propose that Ago1-bound small RNAs target RITS to centromeres via base-pairing interactions with nascent, chromatin-associated non-coding transcripts. Consequently, RITS recruits the RNA-dependent RNA polymerase complex (RDRC) to initiate double-stranded RNA synthesis and siRNA amplification, as well as the cryptic loci regulator complex (CLRC) to facilitate methylation of histone H3K9 (ref. 4). Chp1 reinforces the heterochromatin association of RITS by binding methylated H3K9 with high affinity[5], thereby creating a positive-feedback loop between siRNA biogenesis, RITS localization and H3K9 methylation. Hence, siRNA-programmed RITS acts as a specificity determinant for the recruitment of other RNAi complexes and chromatin-modifying enzymes to centromeres. However, an outstanding question is whether synthetic siRNAs can also function in this context, and thereby be used to trigger *de novo* formation of heterochromatin, particularly outside of centromeric repeats, to stably silence gene expression at will[1].

Small RNAs have crucial roles in endogenous chromatin-mediated processes also in plants, *Caenorhabditis elegans*, *Drosophila melanogaster* and ciliates. Their role in chromatin silencing can also be extended to mammalian cells, although the mechanisms and physiological pathways are less clear[1,6]. Yet, efforts to initiate chromatin modifications *in trans* by using siRNAs have been inherently difficult to achieve in all organisms. In plants, this is because the ability of siRNAs to induce DNA methylation at gene promoters is context-dependent and sensitive to pre-existing chromatin modifications[7]. And although siRNAs have been shown to promote DNA methylation *in trans* on homologous reporter transgenes in tobacco and *Arabidopsis*[8], it is unclear whether this is a general phenomenon for endogenous promoters. In mammalian cells, the introduction of siRNAs or hairpin RNAs has been reported to promote the modification of DNA and histones[9–11]. However, most small RNAs seem to mediate post-transcriptional gene silencing exclusively, and siRNA-mediated silencing of transcription does not necessarily require chromatin modification[12,13]. Consequently, the potential of synthetic siRNAs to trigger long-lasting gene repression in mammalian cells is debated. Similarly, although studies in *S. pombe* have shown that RNA-hairpin-derived siRNAs can promote H3K9 methylation *in trans* at a small number of loci[14,15], it is inefficient, locus-dependent, and the silent state observed is weak and highly unstable[14]. Rather, endogenous protein-coding genes seem to be refractory to siRNA-directed repression *in trans* in wild-type cells (Extended Data Figs 1 and 2). Therefore, it has been proposed that the ability of siRNAs to direct *de novo* formation of heterochromatin *in trans* is under strict control by mechanisms that have thus far remained elusive.

To identify putative suppressors of siRNA-mediated heterochromatin formation, we designed a small-RNA-mediated silencing (sms) forward genetic screen. We constructed a reporter strain (sms0), which expresses an RNA hairpin (ade6-hp) that is complementary to 250 nucleotides of *ade6+* (Fig. 1a and Extended Data Fig. 1). We chose *ade6+* as a reporter because *ade6* mutant cells form red colonies on limiting adenine indicator plates, whereas *ade6+* cells appear white. Although the ade6-hp construct generated siRNAs complementary to *ade6+* messenger RNAs, no red colonies were visible, demonstrating that *ade6+* siRNAs cannot silence the *ade6+* gene *in trans* in sms0 cells (Extended Data Figs 1b and 2). To screen for mutants that would enable *ade6+* siRNAs to act *in trans*, we mutagenized sms0 cells with ethylmethansulfonate (EMS). This revealed five *sms* mutants that are highly susceptible to *de novo* formation of heterochromatin and stable gene silencing by siRNAs that are acting *in trans* (Extended Data Fig. 3).

[1]Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland. [2]University of Basel, Petersplatz 10, 4003 Basel, Switzerland. [3]Swiss Institute of Bioinformatics, Maulbeerstrasse 66, 4058 Basel, Switzerland. [4]Eötvös Loránd University, Faculty of Sciences, Institute of Chemistry, 1/A Pázmány Péter sétány, Budapest 1117, Hungary.
*These authors contributed equally to this work.

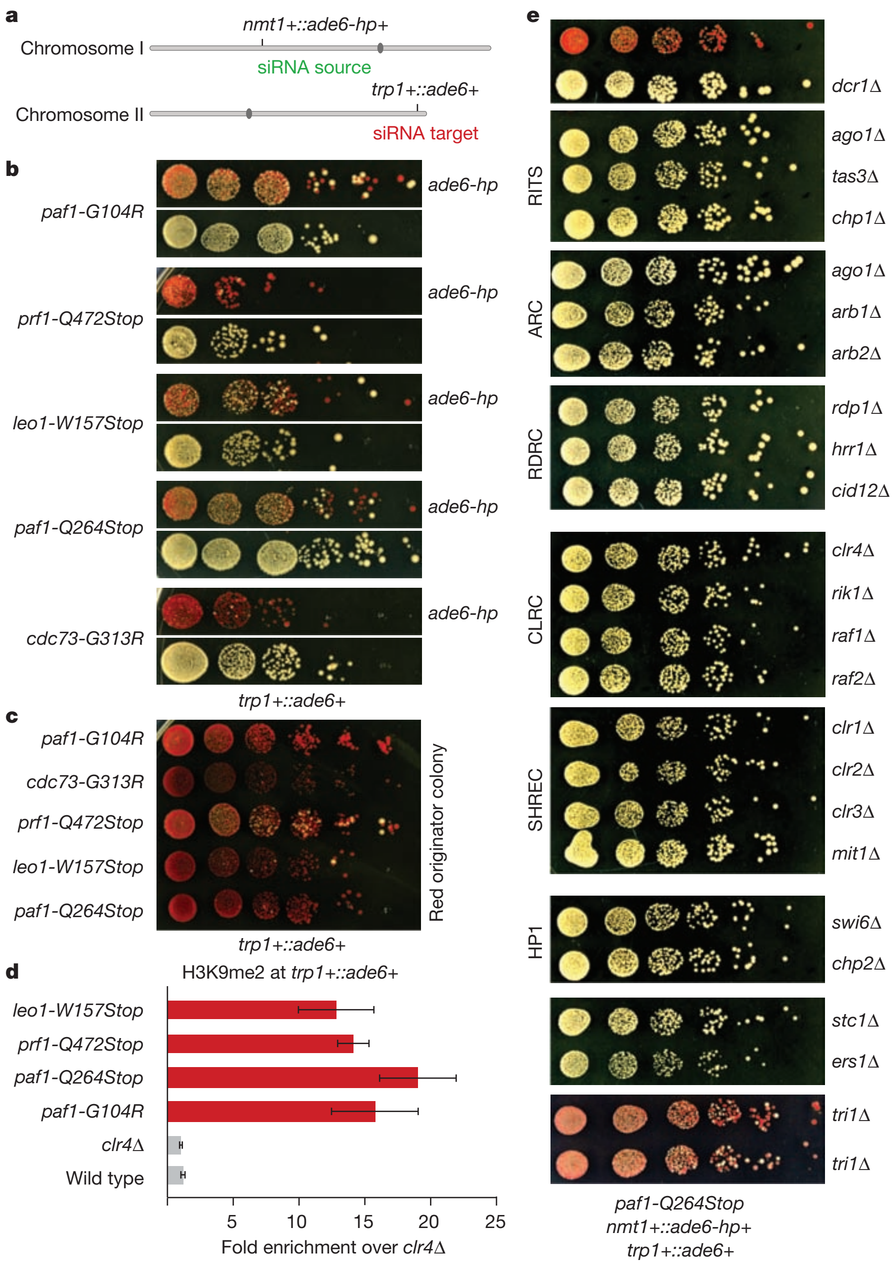


Figure 1 | siRNA-directed *de novo* formation of heterochromatin. a, The ade6-hp RNA producing locus and siRNA target *in trans* in the sms0 strain. **b**, Silencing assay performed with freshly generated Paf1C mutants. **c**, Silencing assay performed with red colonies from **b**. **d**, *ade6+* siRNAs direct the methylation of H3K9 *in trans* at the *trp1+::ade6+* locus in Paf1C mutant cells. Error bars, s.e.m.; $n = 3$ technical replicates. **e**, Gene silencing at the *trp1+::ade6+* locus depends on the same factors as constitutive heterochromatin at centromeric repeats. ARC, Argonaute siRNA chaperone complex.

To map the mutations in *sms* mutants, we re-sequenced the genomes of sms0 and backcrossed *sms* mutants using whole-genome next-generation sequencing. We mapped missense or nonsense mutations in the genes SPBC651.09c, SPAC664.03, SPBC13E7.08c and SPBC17G9.02c (Extended Data Fig. 3), whose homologues in budding yeast encode for protein subunits of the Paf1 complex. We therefore named SPAC664.03, SPBC13E7.08c and SPBC17G9.02c after the *S. cerevisiae* homologues *paf1+*, *leo1+* and *cdc73+*, respectively. SPBC651.09c has already been named as *prf1+* (ref. 16). To validate these as the causative mutations, we reconstituted the candidate point mutations in Paf1, Leo1, Cdc73 and Prf1 in sms0 cells. All five point mutations recapitulated the *sms* mutant phenotype in cells expressing ade6-hp siRNAs (Fig. 1b, c). As expected from the red colour assays, *ade6+* mRNA levels were reduced in all mutant strains. siRNA-mediated *ade6+* silencing was also observed in cells that express a carboxy-terminally 3×Flag-tagged version of the fifth Paf1C subunit Tpr1, which acts as a hypomorphic allele (Extended Data Fig. 4). Therefore, we have identified mutant alleles for the homologues of all five subunits of Paf1C that enable siRNAs to induce gene silencing *in trans*.

We next analysed whether other genes could also be silenced *in trans* in the Paf1C mutants. We first selected the endogenous *ura4+* gene, as this has been shown to be refractory to silencing by siRNAs acting *in trans*[14,15,17]. The *paf1-Q264Stop* mutation was introduced in a strain expressing *ura4+* siRNAs from a *ura4+* hairpin integrated at the *nmt1+* locus[15]. *ura4+* repression was monitored by growing cells on media containing 5-fluoroorotic acid (5-FOA), which is toxic to *ura4+* expressing cells. As expected, *paf1+* cells did not grow on 5-FOA-containing media, indicating that the *ura4+* gene is expressed. However, *paf1-Q264Stop* cells formed colonies on 5-FOA containing media, demonstrating siRNA-directed silencing of the endogenous *ura4+* locus (Extended Data Fig. 5a). Similarly, siRNAs generated at the heterochromatic *ura4+::5BoxB* locus[18] were able to silence a *leu1Δ::ura4+* reporter *in trans* in *paf1-Q264Stop* but not *paf1+* cells (Extended Data Fig. 5b), demonstrating that siRNAs generated from sources other than RNA stem–loop structures also direct *trans*-silencing in *paf1+* mutant cells. Finally, we also observed silencing of the endogenous *ade6+* gene when ade6-hp siRNAs were expressed from the *nmt1+* locus in *paf1-Q264Stop* cells (Extended Data Fig. 5c). In summary, Paf1C mutations

enabled siRNA-directed silencing *in trans* at all euchromatic loci that we tested.

The foregoing results indicated that *de novo* formation of heterochromatin was mediated by *trans*-acting siRNAs. Indeed, Paf1C mutants showed high H3K9 methylation at all *ade6+* siRNA target loci (Fig. 1d and Extended Data Fig. 6a–c), demonstrating that Paf1C prevents *trans*- as well as *cis*-acting siRNAs from directing methylation of H3K9. Further corroborating the formation of bona fide heterochromatin at the *ade6+* target locus, *ade6+* repression was dependent on components of SHREC (Snf2/histone deacetylase (HDAC) repressor complex) and CLRC, as well as the heterochromatin protein 1 (HP1) proteins Swi6 and Chp2, which are known to facilitate constitutive heterochromatin formation at centromeres (Fig. 1e). Finally, the formation of heterochromatin reduced transcriptional activity of the *ade6+* gene as evidenced by reduced H3K36 tri-methylation and RNA polymerase (Pol) II occupancy (Extended Data Fig. 6d, e). From these results we conclude that siRNAs can initiate the formation of heterochromatin and gene silencing, but that this is under strict negative control by Paf1C. This explains previous unsuccessful attempts to induce stable heterochromatin formation *in trans* using synthetic siRNAs.

Consistent with the formation of an epigenetically distinct chromatin domain at the siRNA target loci, cells in a population of freshly generated Paf1C mutants were either fully red or fully white. The latter gradually became red with increasing numbers of mitotic divisions, and once established, the silent state was remarkably stable (Fig. 1b, c). The fact that not all cells in a population of naive Paf1C mutant cells turned red immediately allowed us to determine the frequency of initiation of heterochromatin formation quantitatively. This analysis revealed that silencing in mitotic cells was efficiently established in *leo1-W157Stop* mutant cells, whereas *cdc73-G313R* cells were the least efficient (Fig. 2a). Descendants of a red colony switched to the white phenotype only sporadically in all Paf1C mutants, demonstrating that maintenance of heterochromatin is very robust in these cells (Fig. 2b). Interestingly, siRNA-directed *de novo* formation of heterochromatin was most efficient in meiosis. In 70% of all crosses between a naive *paf1-Q264Stop* mutant (white) and a *paf1+* cell, at least one of two *paf1-Q264Stop* spores had initiated *ade6+* repression (red) (Fig. 2c and Extended Data Fig. 7). We also observed highly efficient propagation of the silent state through meiosis, but only in descendants of spores that inherited the Paf1C mutation (Fig. 2d). Thus, siRNAs are sufficient to initiate the formation of very stable heterochromatin when Paf1C function is impaired.

Notably, assembly of heterochromatin at the *ade6+* target gene was accompanied by the production of novel *ade6+* siRNAs that are not encoded in the ade6-hp and that accumulated to high levels (Fig. 2e). Thus, primary ade6-hp siRNAs trigger the production of highly abundant secondary *ade6+* siRNAs in Paf1C mutants. To test whether continuous production of siRNAs is necessary for sustaining the repressed state, we deleted genes encoding for RNAi factors and found that *ade6+* silencing was completely abolished in all canonical RNAi mutants. Deletion of *tri1+* resulted in moderate derepression of *ade6+* silencing, suggesting a minor contribution of this exonuclease to siRNA-mediated heterochromatin silencing (Fig. 1e). To test whether secondary siRNAs produced at the *ade6+* target locus are sufficient to maintain heterochromatin, we crossed a *trp1+::ade6+ paf1-Q264Stop ade6-hp+* strain (red) with a *trp1+::ade6+ paf1-Q264Stop* (white) strain. These crosses regularly produced spores that gave rise to red cells even in the absence of the *nmt1+::ade6-hp+* allele. The red phenotype was still visible after replica plating, demonstrating that heterochromatin can be maintained in the absence of the primary siRNAs for hundreds of mitotic cell divisions (Fig. 2f). These results demonstrate that siRNAs can induce an epigenetic change in gene expression in meiotic and mitotic cells, and that secondary siRNA production is sufficient to propagate the repressed state for many mitotic cell divisions independently of the primary siRNAs that triggered the epigenetic switch.

The highly conserved Paf1C is well known for promoting RNA Pol II transcription elongation and RNA 3′-end processing (Fig. 3a). Paf1C also governs transcription-coupled histone modifications and has connections to DNA damage repair, cell cycle progression, and other processes[19]. Given this broad function, we assessed the effect of our Paf1C mutations on genome expression. This analysis revealed that *paf1-G104R*, *paf1-Q264Stop*, *prf1-Q472Stop* and *leo1-W157Stop* impair repression of heterochromatin formation, without affecting RNA expression globally (Supplementary Information and Extended Data Fig. 8). This is consistent with our observation that *ade6+* expression is unaffected in Paf1C mutants in the absence of siRNAs (Fig. 1b). We did, however, detect a reduction in H3K36 tri-methylation and an increase in RNA Pol II occupancy on the *ade6+* gene in *paf1-Q264Stop* cells (Fig. 3b, c). This is consistent with the role of Paf1C in promoting transcription, and suggests that decelerated transcription kinetics in Paf1C mutants enables siRNA-directed epigenetic gene silencing. To dissect which of Paf1C's activities are most critical to prevent RNAi-mediated heterochromatin assembly, we interfered with transcription elongation, termination

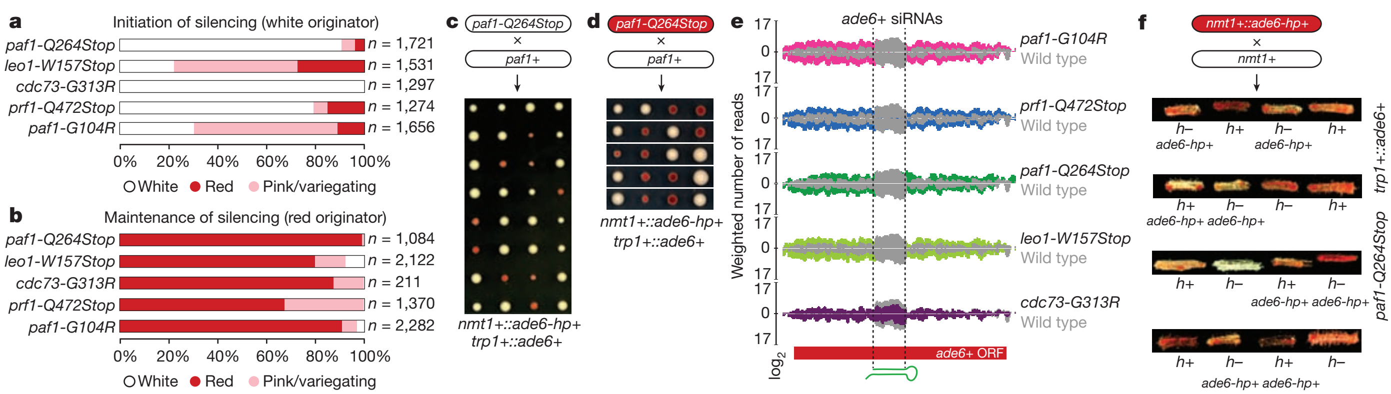


Figure 2 | siRNA-mediated epigenetic gene silencing. **a**, Percentage of naive Paf1C mutant cells that establish heterochromatin within 20–30 mitotic divisions. *n*, number of scored colonies. **b**, Stability of ectopic heterochromatin in mitotic cells. *n*, number of scored colonies. **c**, Initiation of heterochromatin formation during meiosis. Naive *paf1-Q264Stop* cells (white) were crossed with *paf1+* cells. Spore dissection of eight crosses is shown. **d**, Red *paf1-Q264Stop* cells (heterochromatic *ade6+*) were crossed with *paf1+* cells to assess stability of ectopic heterochromatin through meiosis. White descendants are *paf1+*. **e**, siRNA reads mapping to the *ade6+* locus in wild-type (grey) and Paf1C mutant (coloured) strains. Read counts were normalized to library size and are shown in $\log_2$ scale. Dashed lines mark the *ade6+* fragment targeted by the hairpin. **f**, Red *paf1-Q264Stop* cells (heterochromatic *ade6+*) carrying the *ade6+*-targeting hairpin (*ade6-hp+*) were crossed with *paf1-Q264Stop* cells without the hairpin to test hairpin requirement after initiation of silencing. Four spores derived from the cross were struck on yeast extract (YE) plates to assess the silencing phenotype. *h+* and *h−* denote mating types and *ade6-hp+* marks cells carrying the hairpin.

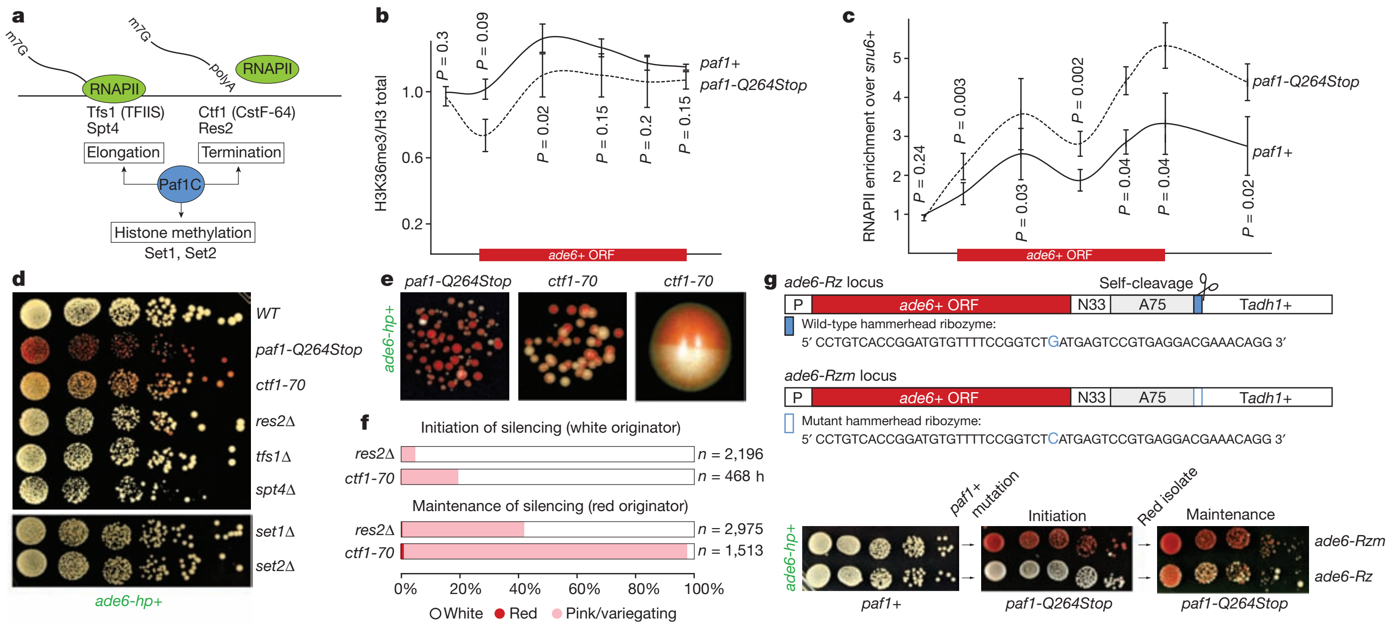


Figure 3 | Mechanism of repression. **a**, Paf1C governs RNA Pol II (RNAPII) transcription elongation, RNA 3′-end processing, and transcription-coupled histone modifications. **b**, **c**, Chromatin immunoprecipitation experiments to assess *ade6+* transcriptional activity in Paf1C mutant cells. H3K36me3 levels were normalized to total H3 levels. *snu6+* is transcribed by RNAPIII and serves as background control. ORF, open-reading frame. Error bars, s.e.m.; $n = 3$ independent biological replicates; *P* values were calculated using one-tailed Student's *t*-test. **d**, **e**, Silencing assays showing that *ade6+* siRNAs can initiate repression of *ade6+* in transcription termination mutants. Note the bi-stable state of repression in *ctf1-70* cells. Cells were grown on YE plates. WT, wild type. **f**, Percentage of naive transcription termination defective cells (white originator) that establish heterochromatin within 20–30 mitotic divisions (initiation) and stability of ectopic heterochromatin in descendants thereof (maintenance). *n*, number of scored colonies. **g**, A 52-nucleotide wild-type or mutant hammerhead ribozyme sequence preceded by a templated polyA(75)-tail was integrated 33 nucleotides downstream of the *ade6+* stop codon (*ade6-Rz* or *ade6-Rzm*, respectively). **h**, Silencing assay showing that *ade6+* siRNAs stably repress *ade6-Rzm* but not *ade6-Rz* in *paf1* mutant cells. Note that *ade6-Rz* produces fully functional mRNA in *paf1+* cells.

or co-transcriptional histone modification directly by mutating genes encoding elongation factors (Tfs1 and Spt4), termination factors (Ctf1 and Res2), or histone methyltransferases (Set1 and Set2)[20,21] (Fig. 3a). We observed siRNA-mediated initiation of *ade6+* silencing in *ctf1-70* and *res2Δ* cells, but not in *tfs1Δ*, *spt4Δ*, *set1Δ* and *set2Δ* cells (Fig. 3d–f), demonstrating that impaired transcription termination but not elongation is sufficient to allow siRNA-directed repression. Notably, although impaired transcription termination in *ctf1-70* and *res2Δ* cells was sufficient to initiate silencing, the silent state was less stable than in *paf1-Q264Stop* mutant cells (Fig. 3e, f). This explains why our screen did not reveal mutations in transcription termination factors.

In *ctf1-70* cells, although RNA Pol II fails to terminate, the nascent RNA is still properly processed and released from the site of transcription[21]. This probably accounts for the less stable silencing in *ctf1-70* cells and suggests that the more severe phenotype of Paf1C mutants is due to the combined effects of impaired termination and nascent transcript release. Therefore, we tested whether artificially releasing the nascent transcript from the site of transcription partially alleviates siRNA-mediated heterochromatin formation in Paf1C mutant cells. To this end, we inserted a 52-nucleotide hammerhead ribozyme (Rz), preceded by a templated polyA (A_{75}) tail, downstream of the *ade6+* open reading frame (*ade6-Rz*) to induce self-cleavage of nascent *ade6+* transcripts (Fig. 3g). Indeed, initiation of silencing at the *ade6-Rz* locus was inefficient and the repressed state was poorly propagated in *paf1-Q264Stop* mutant cells. By contrast, silencing was very effective in cells that contain a single base change in the catalytic site of the ribozyme (*ade6-Rzm*) that abolishes self-cleavage (Fig. 3h). Thus, retaining the nascent transcript on chromatin is critical to stabilize the repressed state.

These results are consistent with a kinetic model for Paf1C function and demonstrate that proper transcription termination is crucial to prevent *de novo* formation of heterochromatin by siRNAs (Extended Data Fig. 9). This is further supported by the recent observation that termination sequences in the 3′ untranslated region of the *ura4+* gene inhibit the ability of siRNAs to promote heterochromatin formation[17] and is reminiscent of enhanced silencing phenotype (*esp*) mutations in *Arabidopsis thaliana*, which are in genes that encode for members of the cleavage polyadenylation specificity factor and cleavage stimulation factor complexes[22]. Importantly, our results show that impairment of both transcription termination and nascent transcript release is imperative to confer stability to the repressed state, although compromised transcription termination is sufficient to initiate the formation of bi-stable heterochromatin by *trans*-acting siRNAs.

Besides Dcr1-dependent siRNAs, Ago1 associates with Dcr1-independent small RNAs referred to as primal RNAs (priRNAs). priRNAs seem to be degradation products of abundant transcripts and could potentially trigger siRNA amplification and uncontrolled heterochromatic gene silencing[23]. Therefore, we speculated that the physiological function of Paf1C is to protect the genome from spurious priRNA-mediated heterochromatin formation. To investigate this we analysed whether Paf1C mutants would disclose genomic regions that could be potentially assembled into facultative heterochromatin by endogenous small RNAs. On the basis of our results, loci at which facultative heterochromatin forms in an RNAi-dependent manner are expected to show reduced RNA expression with a concomitant increase in siRNA production. As expected, the *nmt1+::ade6-hp+*, *trp1+::ade6+* and *ade6-704* loci fulfilled this criteria (Extended Data Fig. 10a). Moreover, we observed repression and siRNA production for genes flanking these loci, indicating spreading of heterochromatin into neighbouring genes, which occurred up to 6 kilobases (kb) up or downstream of the ade6-hp siRNA target sites. Indeed, we observed H3K9 methylation in this region in *leo1Δ* cells specifically (Extended Data Fig. 10b, c). In addition to these regions, we observed siRNA-directed silencing signatures at different, non *ade6+*-linked genomic loci, indicating that Paf1C may indeed function to protect the genome from illegitimate repression of protein coding genes by endogenous priRNAs. However, we did not recover the same sites repeatedly in the different Paf1C mutants

(Supplementary Table 1). This indicates that initiation of silencing at these sites occurred stochastically and that there are no specific sites primed for the formation of facultative heterochromatin in mitotic cells that are grown under standard laboratory conditions. Therefore, we conclude that Paf1C protects protein-coding genes from unwanted long-term silencing that might occur by chance, thereby restraining phenotypic variation and conferring epigenetic robustness to the organism.

In summary, we discovered that synthetic siRNAs are highly effective in directing locus-independent assembly of heterochromatin that can be stably maintained through mitosis and meiosis only when Paf1C activity is impaired. A remarkable observation of our study is that the newly established heterochromatin was inherited for hundreds of cell divisions across generations in Paf1C mutant cells, even in the absence of the primary siRNAs that triggered the assembly of heterochromatin. This phenomenon complies with the classical definition of epigenetics[24] (that is, that it is heritable even in the absence of the initiating signal) and highlights fundamental roles of Paf1C and the RNAi machinery in building up epigenetic memory. This mechanism is also reminiscent of RNA-mediated epigenetic phenomena in higher eukaryotes such as paramutation[25] and RNA-induced epigenetic silencing (RNAe)[26]. RNAe is a phenomenon in which small RNAs of the *C. elegans* Piwi pathway can initiate transgene silencing that is extremely stable across generations even in the absence of the initiating Piwi protein. Yet, not all Piwi pathway RNAs trigger RNAe (ref. 27). Similarly, generation of siRNAs is necessary but not sufficient for paramutation in maize[28]. Thus, Paf1C may also have a regulatory role in paramutation and/or RNAe. Notably, Paf1C is known to help maintain expression of transcription factors required for pluripotency in human and mouse embryonic stem cells and prevent expression of genes involved in lineage specification[29,30], which may also involve small RNAs and chromatin regulation.

The ability to induce long-lasting and sequence specific gene silencing by transient delivery of synthetic siRNAs without changing the underlying DNA sequence will not only enable fundamental research on mechanisms that confer epigenetic memory, but may also open up new avenues in biotechnology and broaden the spectrum of the potential applications of RNAi-based therapeutics. Epigenetic control over gene expression is of particular interest in plant biotechnology, as this would circumvent the generation of genetically modified organisms.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 October 2014; accepted 16 February 2015.
Published online 25 March 2015.

1. Moazed, D. Small RNAs in transcriptional gene silencing and genome defence. *Nature* **457,** 413–420 (2009).
2. Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297,** 1833–1837 (2002).
3. Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303,** 672–676 (2004).
4. Motamedi, M. R. *et al.* Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119,** 789–802 (2004).
5. Schalch, T. *et al.* High-affinity binding of Chp1 chromodomain to K9 methylated histone H3 is required to establish centromeric heterochromatin. *Mol. Cell* **34,** 36–46 (2009).
6. Castel, S. E. & Martienssen, R. A. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nature Rev. Genet.* **14,** 100–112 (2013).
7. Chan, S. W., Zhang, X., Bernatavichute, Y. V. & Jacobsen, S. E. Two-step recruitment of RNA-directed DNA methylation to tandem repeats. *PLoS Biol.* **4,** e363 (2006).
8. Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A. & Matzke, A. J. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* **19,** 5194–5201 (2000).
9. Ting, A. H., Schuebel, K. E., Herman, J. G. & Baylin, S. B. Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. *Nature Genet.* **37,** 906–910 (2005).
10. Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. Small interfering rna-induced transcriptional gene silencing in human cells. *Science* **305,** 1289–1292 (2004).
11. Kim, D. H., Villeneuve, L. M., Morris, K. V. & Rossi, J. J. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nature Struct. Mol. Biol.* **13,** 793–797 (2006).
12. Janowski, B. A. *et al.* Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nature Struct. Mol. Biol.* **13,** 787–792 (2006).
13. Napoli, S., Pastori, C., Magistri, M., Carbone, G. M. & Catapano, C. V. Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. *EMBO J.* **28,** 1708–1719 (2009).
14. Simmer, F. *et al.* Hairpin RNA induces secondary small interfering RNA synthesis and silencing in *trans* in fission yeast. *EMBO Rep.* **11,** 112–118 (2010).
15. Iida, T., Nakayama, J. & Moazed, D. siRNA-mediated heterochromatin establishment requires HP1 and is associated with antisense transcription. *Mol. Cell* **31,** 178–189 (2008).
16. Mbogning, J. *et al.* The PAF complex and Prf1/Rtf1 delineate distinct Cdk9-dependent pathways regulating transcription elongation in fission yeast. *PLoS Genet.* **9,** e1004029 (2013).
17. Yu, R., Jih, G., Iglesias, N. & Moazed, D. Determinants of heterochromatic siRNA biogenesis and function. *Mol. Cell* **53,** 262–276 (2014).
18. Bühler, M., Verdel, A. & Moazed, D. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125,** 873–886 (2006).
19. Tomson, B. N. & Arndt, K. M. The many roles of the conserved eukaryotic Paf1 complex in regulating transcription, histone modifications, and disease states. *Biochim. Biophys. Acta* **1829,** 116–126 (2013).
20. Cheung, A. C. & Cramer, P. A movie of RNA polymerase II transcription. *Cell* **149,** 1431–1437 (2012).
21. Aranda, A. & Proudfoot, N. Transcriptional termination factors for RNA polymerase II in yeast. *Mol. Cell* **7,** 1003–1011 (2001).
22. Herr, A. J., Molnar, A., Jones, A. & Baulcombe, D. C. Defective RNA processing enhances RNA silencing and influences flowering of *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **103,** 14994–15001 (2006).
23. Halic, M. & Moazed, D. Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. *Cell* **140,** 504–516 (2010).
24. Ptashne, M. On the use of the word 'epigenetic'. *Curr. Biol.* **17,** R233–R236 (2007).
25. Chandler, V. L. Paramutation: from maize to mice. *Cell* **128,** 641–645 (2007).
26. Luteijn, M. J. & Ketting, R. F. PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nature Rev. Genet.* **14,** 523–534 (2013).
27. Luteijn, M. J. *et al.* Extremely stable Piwi-induced gene silencing in *Caenorhabditis elegans*. *EMBO J.* **31,** 3422–3430 (2012).
28. Chandler, V. L. Paramutation's properties and puzzles. *Science* **330,** 628–629 (2010).
29. Ding, L. *et al.* A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell* **4,** 403–415 (2009).
30. Ponnusamy, M. P. *et al.* RNA polymerase II associated factor 1/PD2 maintains self-renewal by its interaction with Oct3/4 in mouse embryonic stem cells. *Stem Cells* **27,** 3001–3011 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Iida for providing the plasmid encoding the ade6-hp construct, N. Laschet and R. Tsuji for technical assistance, S. Thiry for hybridizing tiling arrays, K. Jacobeit and S. Dessus-Babus for small RNA sequencing, T. Roloff for archiving data sets, M. Kirschmann for developing the Matlab script for colony counting, and A. Tuck for comments on the manuscript. This work was supported by funds from the Swiss National Science Foundation, the European Research Council, and the Boehringer Ingelheim Fonds. The Friedrich Miescher Institute for Biomedical Research is supported by the Novartis Research Foundation.

Author Contributions Y.S., K.M.K., V.F. and J.B. generated strains and performed experiments; Y.S. performed the sms screen; the genome-wide small RNA and gene expression data were analysed by K.M.K.; M.B.S. designed and performed the computational analysis of the mutant genome resequencing data; M.B. designed experiments and prepared the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Genome-wide data sets are deposited at GEO under the accession number GSE59171. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B. (marc.buehler@fmi.ch).

METHODS

Strains and plasmids. Fission yeast strains were grown at 30 °C in YES medium. All strains were constructed following a PCR-based protocol[31] or by standard mating and sporulation. Plasmids and strains generated in this study are shown in Supplementary Tables 2 and 3.

EMS mutagenesis, hit selection and backcrossing. Exponentially growing *sms0* (SPB464) cells were washed and resuspended in 50 mM K-phosphate buffer (pH 7.0) and treated with EMS (final concentration 2.5%) for 150 min. An equal volume of freshly prepared 10% sodium thiosulfate was then added. Cells were washed with water and subsequently resuspended in YES. EMS treatment resulted in ~50% cell viability. To screen for mutants in which *ade6+* expression was silenced, cells were spread on YE plates. About 350,000 colonies were examined and pink colonies were selected for further evaluation. Positive hits were backcrossed four times with the parental strains SPB464 or SPB1788, depending on mating type.

Silencing assays. To assess *ura4+* expression, serial tenfold dilutions of the respective strains were plated on PMGc (non-selective, NS) or on PMGc plates containing 2 mg ml^{-1} 5-FOA. To assess *ade6+* expression, serial tenfold dilutions of the respective strains were plated on YES and YE plates.

Assessment of initiation versus maintenance of ectopic heterochromatin formation. Mutant strains were seeded on YE plates and single-cell-derived red or white colonies were selected. Colonies were resuspended in water and 100–500 cells were seeded on YE plates, which were then incubated at 30 °C for 3 days. Images of the plates were acquired after one night at 4 °C and colonies were counted automatically using Matlab (The MathWorks) and ImageJ Software (National Institutes of Health).

RNA isolation and cDNA synthesis. RNA isolation and cDNA synthesis was performed as described previously[32].

Quantitative real-time PCR. Real-time PCR on cDNA samples and ChIP DNA was performed as described[33] using a Bio-Rad CFX96 Real-Time System using SsoAdvanced SYBR Green supermix (Bio-Rad). Primer sequences are given in Supplementary Table 4.

Chromatin immunoprecipitation. Chromatin immunoprecipitation (ChIP) experiments were performed as previously described[33] with minor modifications. In brief, *S. pombe* cells were fixed with 1% formaldehyde for 15 min and then lysed in buffer containing 50 mM HEPES/KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 1 mM phenylmethylsulfonyl fluoride (PMSF) and protease inhibitor cocktail. Chromatin was sheared with a Bioruptor (Diagenode). The following antibodies were used in this study: histone H3K9me2-specific mouse monoclonal antibody from Wako (MABI0307), histone H3-specific rabbit polyclonal antibody from Abcam (ab1791), histone H3K36me3-specific rabbit polyclonal antibody from Abcam (ab9050), and RNA polymerase II mouse monoclonal antibody from Covance (8WG16).

Small RNA sequencing. Total RNA was isolated from exponentially growing cells using the hot phenol method[34]. The RNA was fractionated using RNeasy Midi columns (Qiagen) following the RNA cleanup protocol provided by the manufacturer. The flow-through fraction was precipitated ('small RNA' fraction). Aliquots (25 µg) of the small RNA fraction were separated by 17.5% PAGE and the 18–28-nucleotide population purified. Libraries were prepared using the Illumina TruSeqTM small RNA preparation protocol (RS-930-1012). The 145–160-nucleotide population was isolated and the library sequenced on an Illumina HiSeq2000. Small RNA reads were aligned as described previously[32] with two mismatches allowed.

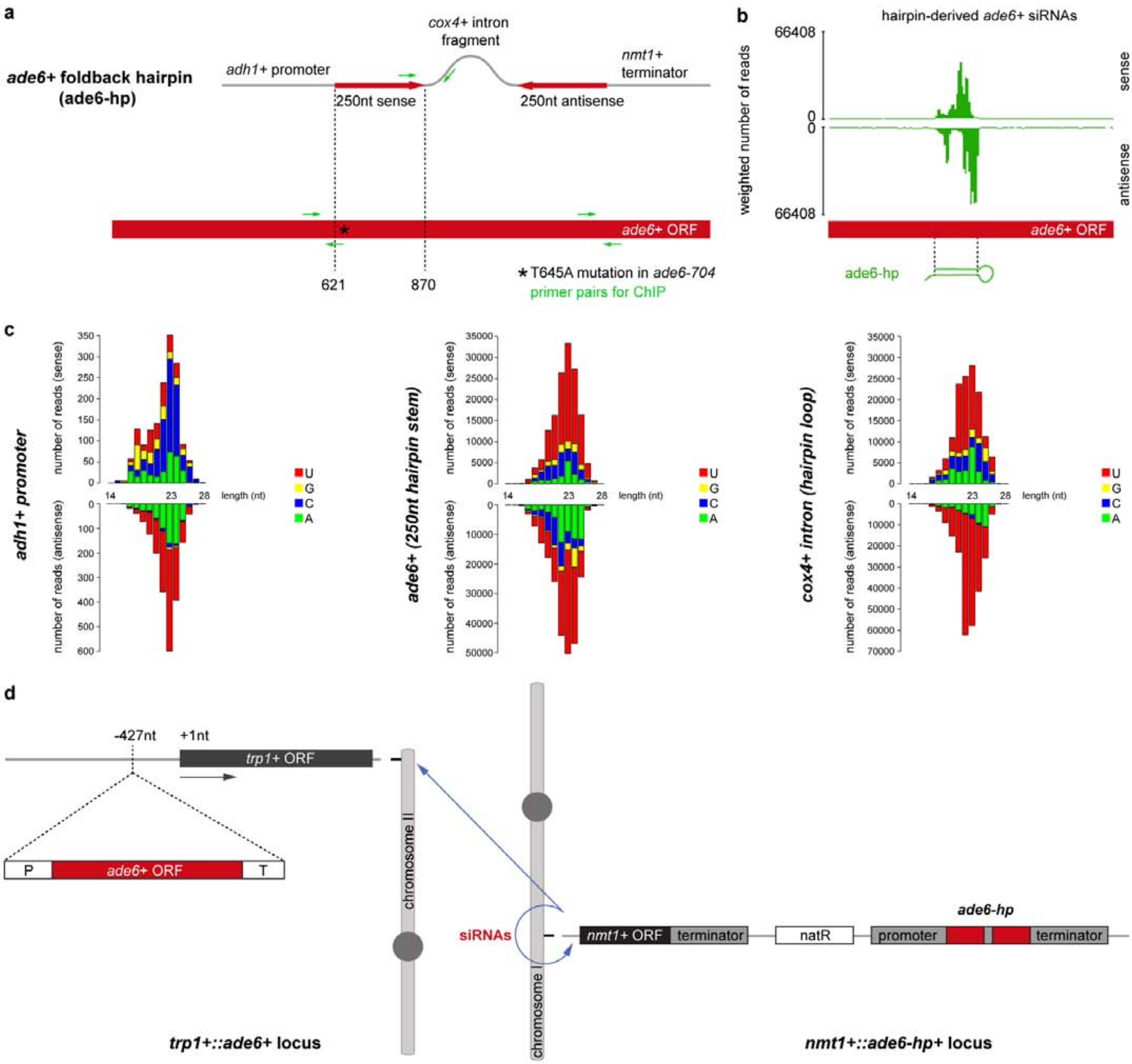
Whole-genome sequencing. Cells from an overnight culture were collected, washed once with water and flash frozen in liquid nitrogen. Cells were spheroplasted in spheroplast buffer (1.2 M sorbitol, 100 mM $KHPO_4$, pH 7.5, 0.5 mg ml^{-1} zymolyase (Zymo Research), 1 mg ml^{-1} lysing enzyme from *Trichoderma harzianum* (Sigma)). Genomic DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen). Barcoded genomic DNA libraries for Illumina next-generation sequencing were prepared from 50 ng genomic DNA using the Nextera DNA Sample Preparation Kit (Illumina). Libraries were pooled equimolarly and sequenced on one lane of a HiSeq2000 machine (Illumina). Basecalling was done with RTA 1.13.48 (Illumina) software and for the demultiplexing CASAVA_v1.8.0 (Illumina) was used. For each strain, between 8.7 and 25.5 million (mean of 14.2 million) 50-nucleotide reads were generated and aligned to the *Schizosaccharomyces pombe* 972h- genome assembly (obtained on 17 September 2008 from http://www.broad.mit.edu/annotation/genome/schizosaccharomyces_group/MultiDownloads.html) using 'bwa' (ref. 35, version 0.7.4) with default parameters, but only retaining single-hit alignments ('bwa samse -n 1' and selecting alignments with 'X0:i:1'), resulting in a genome coverage between 26 and 85-fold (mean of 44-fold). The alignments were converted to BAM format, sorted and indexed using 'samtools' (ref. 36, version 0.1.19). Potential PCR duplicates were removed using 'MarkDuplicates' from 'Picards' (http://picard.sourceforge.net/, version 1.92). Sequence variants were identified using GATK (ref. 37, version 2.5.2) indel realignment and base quality score recalibration using a set of high confidence variants identified in an initial step as known variants, followed by single nucleotide polymorphism (SNP) and INDEL discovery and genotyping for each individual strain using standard hard filtering parameters, resulting in a total of 270–274 sequence variations (mean of 280) in each strain compared to the reference genome (406 unique variations in total over all strains). Finally, variations were filtered to retain only high quality single nucleotide variations (QUAL $\geq$ 50) of EMS type (G|C to A|T) with an alellic balance $\geq$ 0.9 (homozygous) that were not also identified in the parental strain (sms0), reducing the number of variations per strain to a number between 2 and 8 (mean of 4.6).

Expression profiling. RNA was isolated from cells collected at an attenuance (D) of 600 nm of 0.5 ($D_{600\,\mathrm{nm}} = 0.5$) using the hot phenol method[34]. The isolated RNA was processed according to the GeneChip Whole Transcript Double-Stranded Target Assay Manual from Affymetrix using the GeneChip *S. pombe* Tiling 1.0FR. All tiling arrays were processed in R[38] using bioconductor[39] and the packages tilingArray[40] and preprocessCore. The arrays were RMA background-corrected, quantile-normalized, and $\log_2$-transformed on the oligonucleotide level using the following command: expr <−$\log_2$(normalize.quantiles(rma.background.correct(exprs(readCel2eSet (filenames,rotated = TRUE))))). Oligonucleotide coordinates were intersected with the genome annotation and used to calculate average expression levels for individual genomic features (excluding those with <10 oligonucleotides) as well as broader annotation categories. In the latter case, multimapping oligonucleotides were counted only once per category (avoiding multiple counts from the same oligonucleotide).

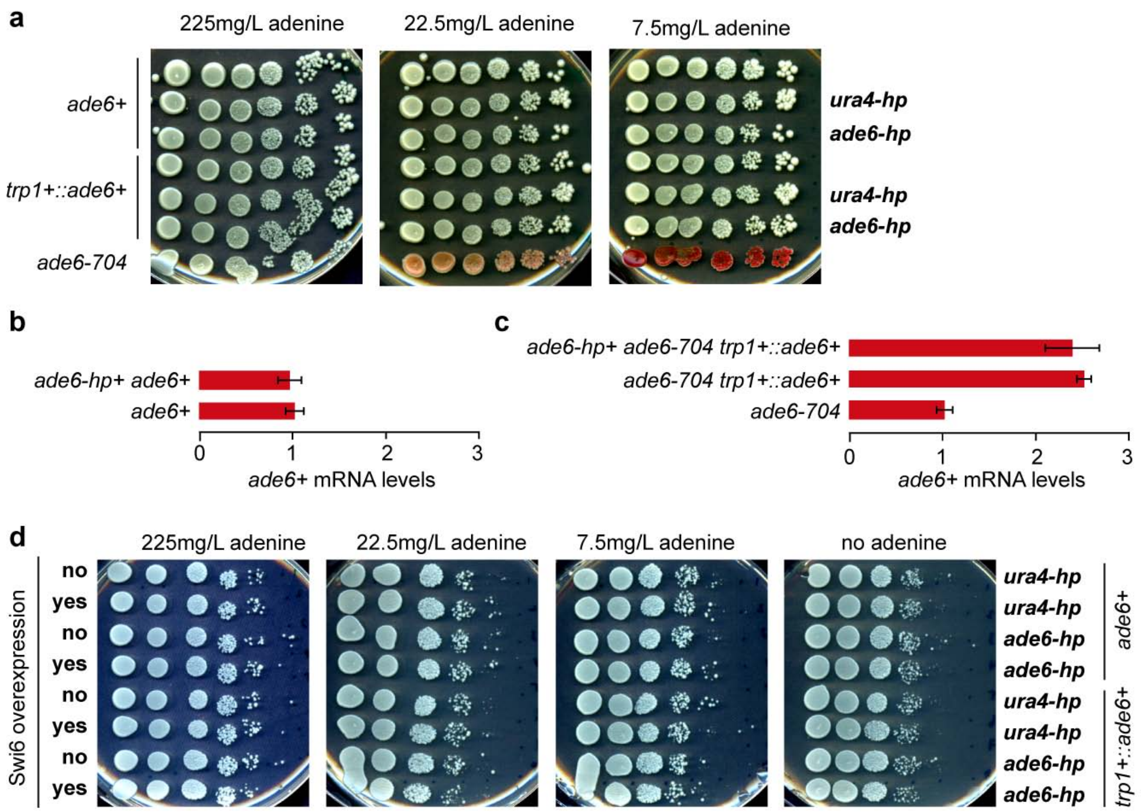
Gene nomenclature. The proteins PAF1p, CDC73p, RTF1p, LEO1p and CTR9p form a stable complex in *S. cerevisiae* (Paf1C). The systematic IDs of the genes encoding the *S. pombe* homologues of these proteins are SPAC664.03, SPBC17G9.02c, SPBC651.09c, SPBC13E7.08c and SPAC27D7.14c, respectively. The *CTR9* homologue SPAC27D7.14c is currently annotated as Tpr1. The RTF1 homologue SPBC651.09c is currently annotated as PAF-related factor 1 (*prf1+*), because *rtf1+* is already used for an unrelated gene (SPAC22F8.07c). Therefore, we refer to SPAC664.03, SPBC17G9.02c, SPBC651.09c, SPBC13E7.08c and SPAC27D7.14c as *paf1+*, *cdc73+*, *prf1+*, *leo1+* and *tpr1+*, respectively, in this paper.

Statistics. A one-tailed Student's t-test was used, with $P < 0.05$ as the significance level. No statistical methods were used to predetermine sample size.

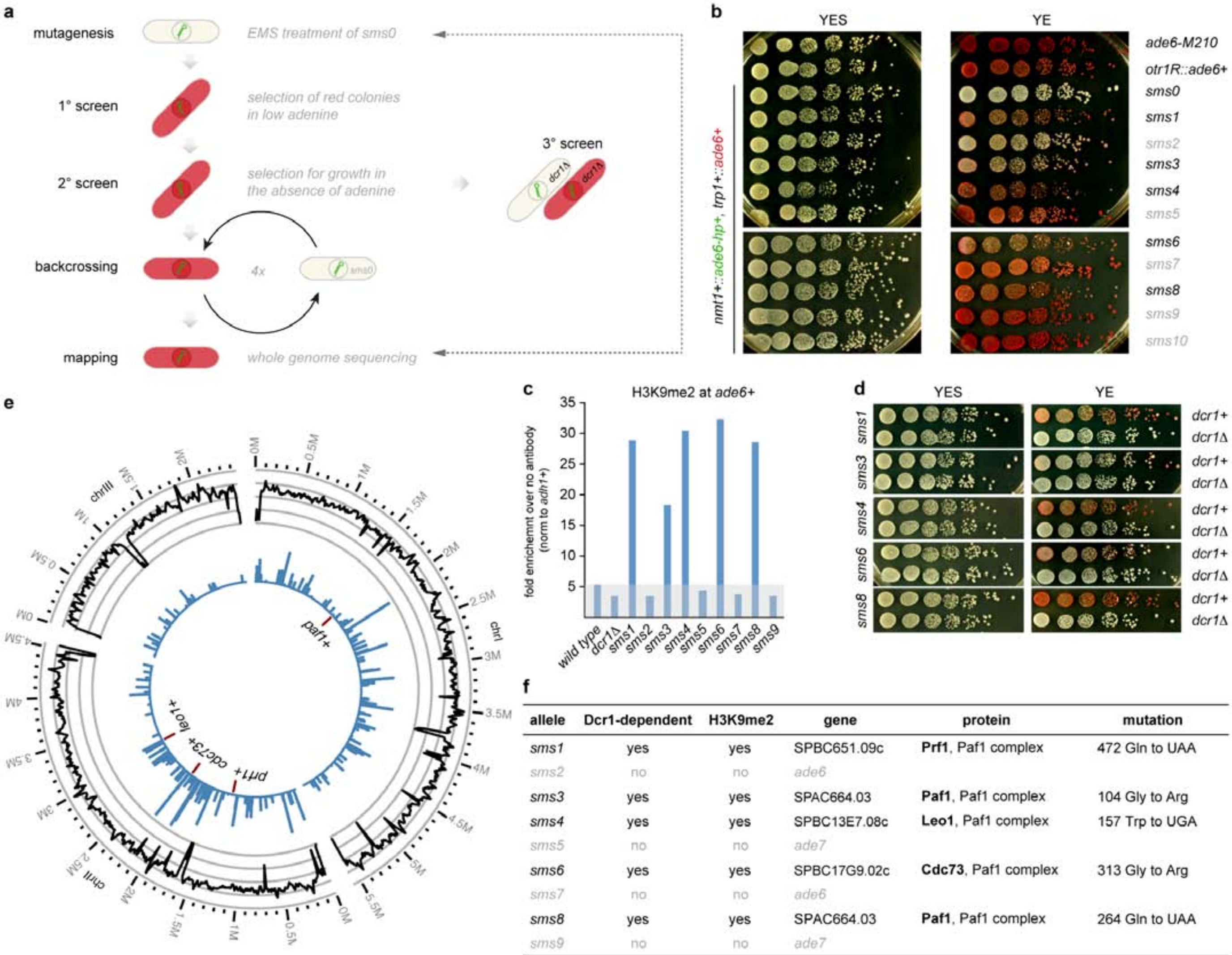
31. Bähler, J. *et al.* Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14,** 943–951 (1998).
32. Emmerth, S. *et al.* Nuclear retention of fission yeast dicer is a prerequisite for RNAi-mediated heterochromatin assembly. *Dev. Cell* **18,** 102–113 (2010).
33. Keller, C. *et al.* HP1(Swi6) mediates the recognition and destruction of heterochromatic RNA transcripts. *Mol. Cell* **47,** 215–227 (2012).
34. Leeds, P., Peltz, S. W., Jacobson, A. & Culbertson, M. R. The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev.* **5,** 2303–2314 (1991).
35. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
36. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
37. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43,** 491–498 (2011).
38. Ihaka, R. & Gentleman, R. R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* **5,** 299–314 (1996).
39. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5,** R80 (2004).
40. Huber, W., Toedling, J. & Steinmetz, L. M. Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics* **22,** 1963–1970 (2006).
41. Bühler, M., Haas, W., Gygi, S. P. & Moazed, D. RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell* **129,** 707–721 (2007).



Extended Data Figure 1 | Design of the *ade6+* RNA hairpin (ade6-hp) construct that expresses abundant sense and antisense (primary) siRNAs. **a**, The RNA stem–loop construct consists of a 250-nucleotide-long *ade6+* fragment, followed by a *cox4+* intronic sequence and the reverse complement of the *ade6+* fragment. The promoter sequence of the *adh1+* gene drives expression of the RNA hairpin. Transcription of the construct is terminated by the termination signals of the *nmt1+* gene. The construct was provided by T. Iida. **b**, **c**, Small RNA sequencing revealed that the RNA stem is converted into sense and antisense siRNAs covering the 250-nucleotide stretch from the *ade6+* open reading frame (nucleotides 621–870). Furthermore, sense and antisense siRNAs mapping to the *cox4+* intronic and *adh1+* promoter sequences are also generated when this construct is expressed in wild-type cells. ORF, open reading frame. Asterisk denotes the point mutation (Thr645Ala) in the *ade6-704* loss of function allele. Green arrows indicate forward and reverse primers that were used for PCR in ChIP experiments. **d**, Schematic diagram depicting origin and target(s) of synthetic ade6-hp siRNAs. The ade6-hp expression cassette (**a**) was inserted into the *nmt1+* locus on chromosome I by homologous recombination. The ade6-hp-containing plasmid was linearized with *PmlI*, which cuts in the middle of the *nmt1+* terminator sequence, and transformed into *ade6-704* cells. Thereby, the ade6-hp construct was inserted downstream of the *nmt1+* gene. The nourseothricin (Nat)-resistance cassette linked to the ade6-hp construct allowed selection of positive transformants. It also allows assessment of spreading of repressive heterochromatin that is nucleated by the ade6-hp siRNAs *in cis* (see Extended Data Fig. 7b). A wild-type copy of the *ade6+* gene was inserted upstream of the *trp1+* gene on chromosome II by homologous recombination. Because the endogenous *ade6-704* allele is non-functional, positive transformants could be selected by growth in the absence of adenine. In Paf1C mutant cells, ade6-hp-derived siRNAs either act *in cis* to assemble heterochromatin at the *nmt1+* locus (chromosome I), or *in trans* to direct the formation of heterochromatin at the *trp1+::ade6+* (chromosome II) and *ade6-704* (chromosome III) loci.

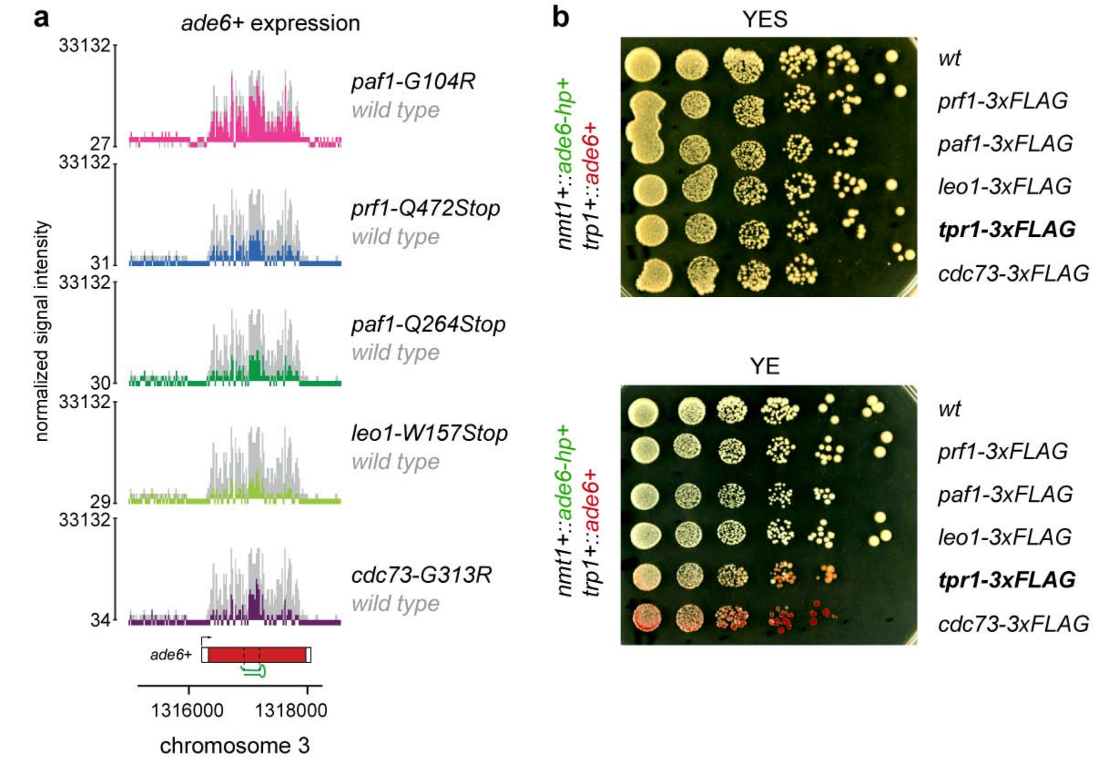


Extended Data Figure 2 | Silencing assays demonstrating the inability of synthetic siRNAs to act *in trans* in Paf1C wild-type cells. **a**, *ade6+* silencing assays were performed with cells expressing synthetic ade6-hp siRNAs, ura4-hp siRNAs or no siRNAs. The ability of ade6-hp siRNAs to silence either the endogenous *ade6+* gene or the *trp1+::ade6+* reporter gene was assessed at different adenine concentrations. *ade6-704* cells were used as positive control. **b**, **c**, *ade6+* mRNA levels were determined by quantitative RT–PCR and normalized to *act1+* mRNA. One representative biological replicate is shown. Error bars, s.d. **d**, *ade6+* silencing assays demonstrating that neither the endogenous *ade6+* gene nor the *trp1+::ade6+* reporter gene becomes repressed by *trans-acting* ade6-hp siRNAs, even upon overexpression of the heterochromatin protein Swi6.

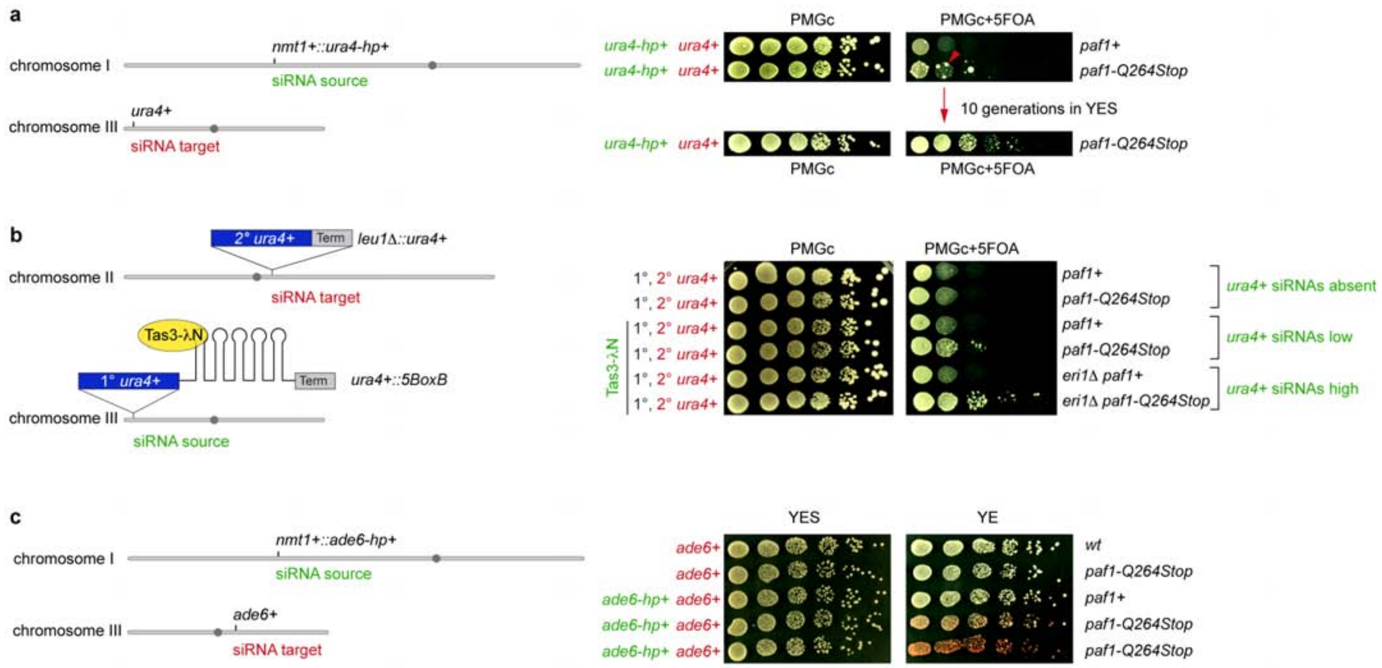


allele	Dcr1-dependent	H3K9me2	gene	protein	mutation
sms1	yes	yes	SPBC651.09c	**Prf1**, Paf1 complex	472 Gln to UAA
sms2	no	no	*ade6*		
sms3	yes	yes	SPAC664.03	**Paf1**, Paf1 complex	104 Gly to Arg
sms4	yes	yes	SPBC13E7.08c	**Leo1**, Paf1 complex	157 Trp to UGA
sms5	no	no	*ade7*		
sms6	yes	yes	SPBC17G9.02c	**Cdc73**, Paf1 complex	313 Gly to Arg
sms7	no	no	*ade6*		
sms8	yes	yes	SPAC664.03	**Paf1**, Paf1 complex	264 Gln to UAA
sms9	no	no	*ade7*		

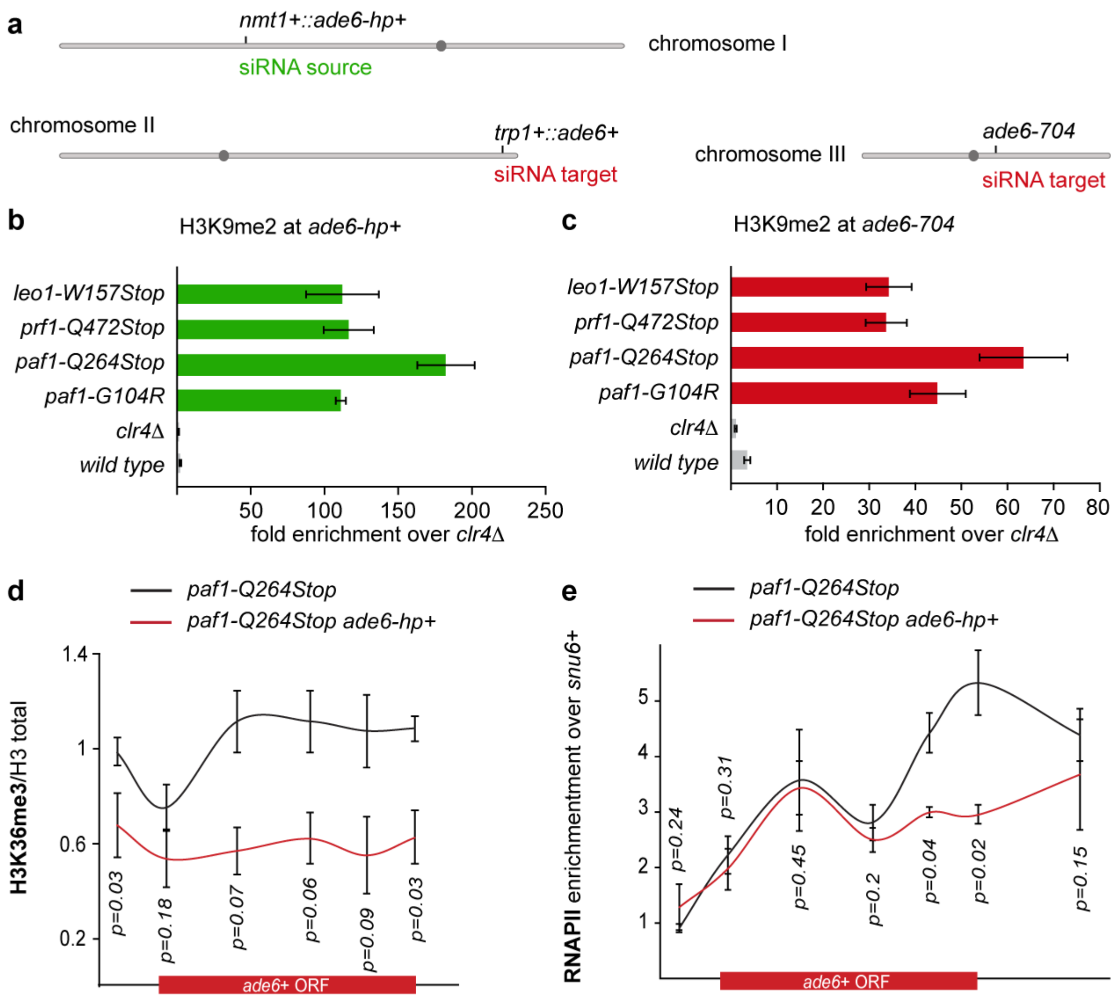
Extended Data Figure 3 | Sms forward genetic screen identifies five true positive hits that enable siRNAs to methylate H3K9 at the *ade6+* gene *in trans*. **a**, Workflow of the EMS mutagenesis screen. We mutagenized sms0 cells, which express abundant siRNAs complementary to the *ade6+* gene (indicated by green hairpin), with EMS (primary screen). Subsequently, we tested the positive red colonies for growth in the absence of adenine to select against loss-of-function mutations in the adenine biosynthesis pathway (secondary screen). In hits that remained positive after the secondary screen, *dcr1+* was deleted to identify truly siRNA-dependent hits (tertiary screen). For mapping of causative mutations by whole-genome next-generation sequencing, positive hits were backcrossed four times. **b**, *sms1-10* mutants show the red *ade6+* silencing phenotype on YE plates, which segregated through four successive backcrosses for all 10 mutants. The *ade6-M210* loss-of-function allele and *ade6+* inserted within centromeric heterochromatin (*otr1R::ade6+*) serve as positive controls. **c**, ChIP experiment demonstrating methylation of H3K9 at the *ade6+* target loci in *sms1*, *3*, *4*, *6* and *8*. One representative biological replicate is shown. **d**, *ade6+* silencing in *sms1*, *3*, *4*, *6* and *8* is Dcr1-dependent. **e**, Resequencing of EMS-mutagenized *S. pombe* strains. From outside to inside, the tracks show the genomic location, the average coverage per window of 10 kb (black line, scale from 0 to 30), the number of sequence variations identified before filtering in all strains per window of 10 kb (blue bars, scale from 0 to 90) and the five mutations that passed the filtering and overlapped with Paf1C genes (red lines, the two mutations in Paf1 are too close to be resolved individually). **f**, Table lists mutations mapped by whole-genome sequencing. In Dcr1-dependent mutants, we mapped mutations in the genes SPBC651.09c, SPAC664.03, SPBC13E7.08c and SPBC17G9.02c whose homologues in budding yeast encode for protein subunits of the Paf1 complex.



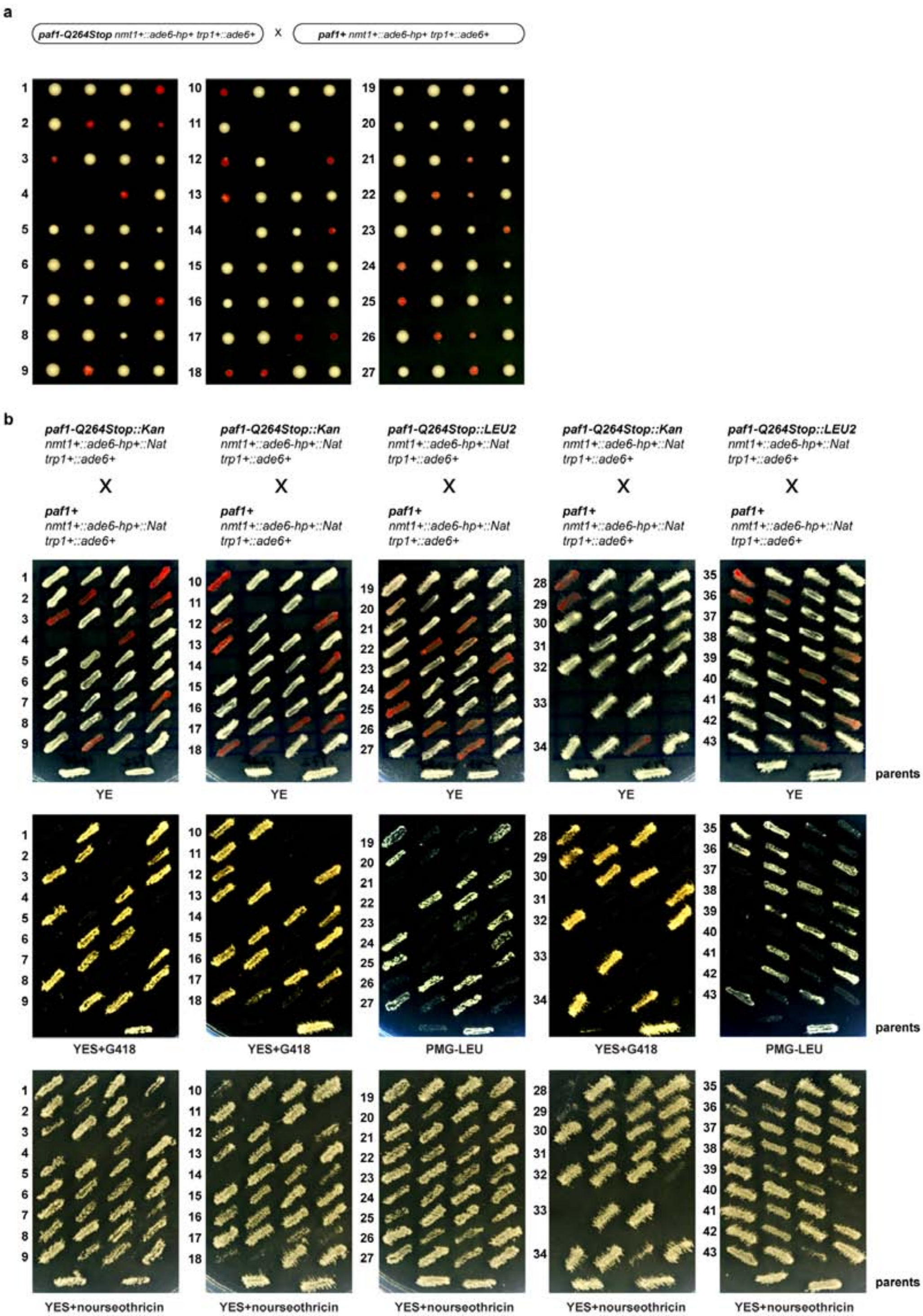
Extended Data Figure 4 | Mutant alleles for the homologues of all five subunits of Paf1C enable siRNAs to induce gene silencing *in trans*. **a**, *ade6+* siRNAs reduce *ade6+* mRNA levels in all Paf1C mutant strains identified in this study. Whole-genome tiling arrays were used to assess gene expression in the mutant cells indicated. *y* axis is in linear scale. **b**, C-terminally tagged Tpr1 and Cdc73 are hypomorphic. Full deletions of the *tpr1+* and *cdc73+* genes cause retarded growth phenotypes (Extended Data Fig. 8c). By contrast, *tpr1-3xFLAG* and *cdc73-3xFLAG* grow normally, and display ade6-hp siRNA-mediated repression of the *ade6+* gene.



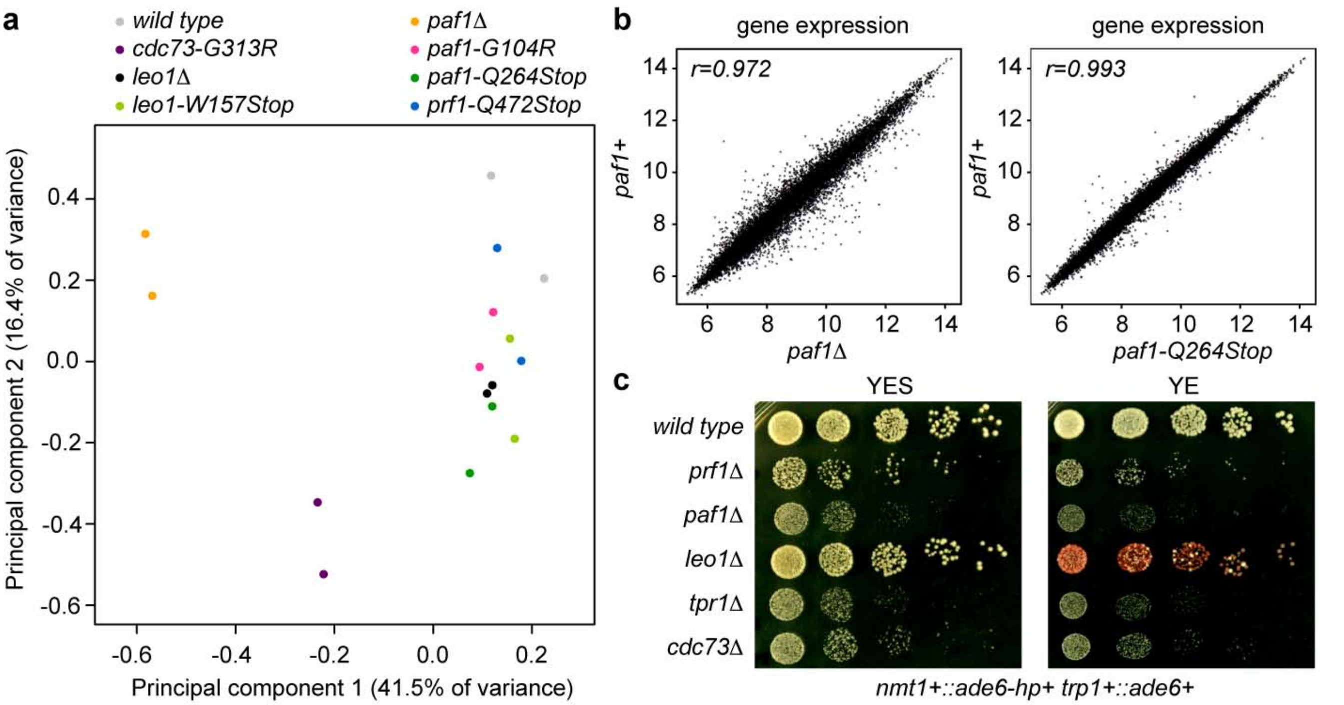
Extended Data Figure 5 | Expression of synthetic siRNAs in *paf1-Q264Stop* cells is sufficient to trigger stable repression of protein coding genes *in trans*. **a**, Left, the *paf1-Q264Stop* mutation was introduced into cells that express synthetic ura4-hp siRNAs[15]. Right, wild-type (*paf1+*) and *paf1-Q264Stop* were grown in the presence or absence of 5-FOA. Red arrow indicates *paf1-Q264Stop* colonies growing on FOA-containing medium. Note that these colonies could be propagated in non-selective medium without losing the repressed state. **b**, In *S. pombe*, artificial tethering of the RITS complex to mRNA expressed from the endogenous *ura4+* locus using the phage λN protein results in *de novo* generation of *ura4+* siRNAs. These siRNAs load onto RITS and are necessary to establish heterochromatin at the *ura4+* locus *in cis*. However, like ura4-hp siRNAs, they are incapable of triggering the repression of a second *ura4+* locus *in trans*[18]. To test whether *ura4+* siRNAs produced as a result of Tas3λN tethering to *ura4+::5BoxB* mRNA (chromosome III) can act *in trans* to silence a second *ura4+* allele (*leu1Δ::ura4+*, chromosome II), *paf1+* was mutated and *ura4+* repression was assessed by FOA silencing assays. Whereas 5-FOA was toxic to both *paf1+* and *paf1-Q264Stop* cells in the absence of *ura4+* siRNAs (Tas3 not fused to λN), FOA-resistant colonies appeared upon Tas3-λN tethering, demonstrating that siRNAs generated from the *ura4+::5BoxB* locus can initiate repression of the second *ura4+* copy expressed from the *leu1+* locus. Notably, siRNA-mediated *ura4+* repression *in trans* was more pronounced in the absence of the RNase Eri1. We have previously shown that the levels of *ura4+::5BoxB*-derived siRNA are higher in *eri1Δ* cells[41]. We note that *trans*-silencing of the second *ura4+* allele occasionally occurs in *paf1+* cells in the absence of Eri1 (ref. 18). However, in contrast to *paf1-Q264Stop* cells, the repressed state of *ura4+* is not stably propagated. Hairpin symbols downstream of the *ura4+* ORF denote BoxB sequences. They form stem–loop structures when transcribed and are bound by the λN protein. **c**, *ade6+* silencing assay demonstrating that also the endogenous *ade6+* gene is repressed if ade6-hp siRNAs are expressed from the *nmt1+* locus in *paf1-Q264Stop* cells. Silencing assay was performed with two freshly generated (naive) *paf1-Q264Stop* mutant strains. A few white colonies in which heterochromatin has not yet formed are discernable. Such white colonies were picked to determine heterochromatin initiation frequencies shown in Fig. 2.



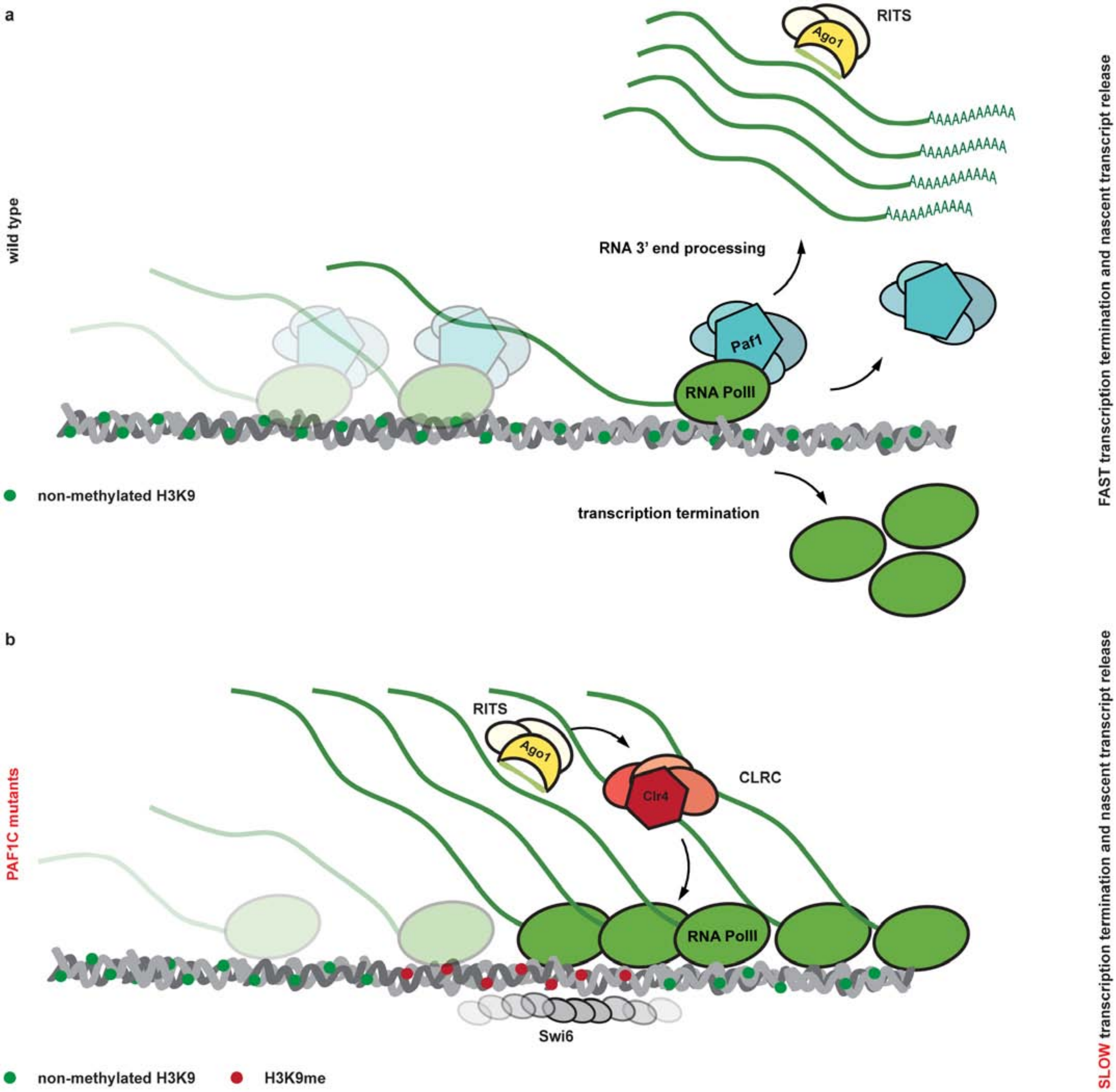
Extended Data Figure 6 | *ade6+* siRNAs trigger *de novo* methylation of H3K9 at homologous *ade6+* sequences *in cis* and *in trans*. **a**, ade6-hp RNA producing locus and siRNA target loci *in trans* in the sms0 strain. *ade6-704* is a loss-of-function allele of the endogenous *ade6+* gene and serves as a positive control in the silencing assays. **b**, **c**, *ade6+* siRNAs direct the methylation of H3K9 at *ade6* targets *in cis* (green) and *in trans* (red) in Paf1C mutant cells. H3K9me2 for *trp1+::ade6+* is shown in Fig. 1d. Quantitative PCR was performed with locus-specific primers. Error bars, s.e.m.; $n = 3$ technical replicates. **d**, **e**, ChIP experiments to assess *ade6+* transcriptional activity. H3K36me3 levels were normalized to total H3 levels. *snu6+* is transcribed by RNAPIII and serves as background control. Error bars, s.e.m.; $n = 3$ independent biological replicates; P values were calculated using the one-tailed Student's t-test.



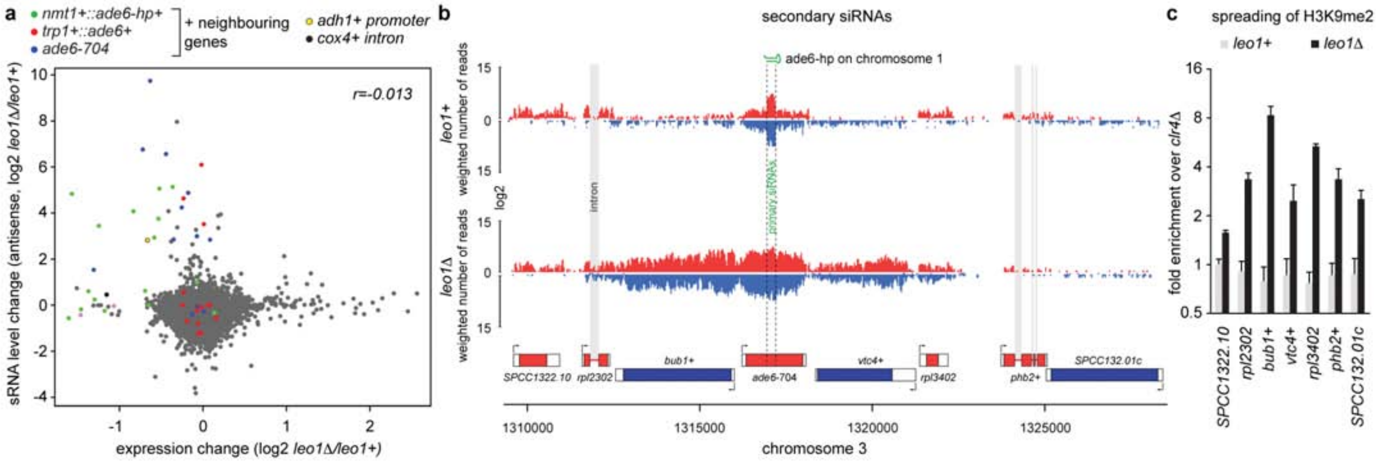
Extended Data Figure 7 | Pronounced siRNA-directed heterochromatin formation *in trans* during meiosis. **a**, White (naive) cells that had not yet established heterochromatin at the *trp1+::ade6+* locus were isolated from populations of freshly generated *paf1-Q264Stop* strains and crossed with *paf1+* cells. Both mating partners expressed ade6-hp siRNAs and contained the same *trp1+::ade6+* reporter. Spores were dissected on YE plates and incubated for 3–4 days at 30 °C. Note the non-Mendelian inheritance pattern of the parental white phenotype and the high incidence of heterochromatin formation (red phenotype) in *paf1-Q264Stop* cells after meiosis. **b**, Spores from 43 tetrads were dissected in total. Colonies formed by the individual spores (**a**) were then struck on YE plates and incubated for 3–4 days at 30 °C, followed by replica-plating onto YES-G418 and YES+nourseothricin (Nat) plates for genotyping. Thus, the cells visible on the YE plates have gone through roughly 50–80 mitotic divisions after mating and sporulation. This analysis shows that *de novo* formation of heterochromatin by *trans*-acting siRNAs during meiosis occurs more frequently than in mitosis. However, once established, heterochromatin is remarkably stable in mitotic cells (see also Fig. 2). Notably, growth of some *paf1-Q264Stop* descendants was reduced on YES+Nat plates, demonstrating spreading of heterochromatin into the neighbouring Nat-resistance cassette that marks the *nmt1+::ade6-hp+* locus (see Extended Data Fig. 1). Note that genes repressed by heterochromatin can be derepressed under strong negative selection. Thus, this observation indicates extraordinary repressive activity of the heterochromatin that forms *in cis* at the ade6-hp siRNA-producing locus. Finally, *paf1+* cells (no growth on YES-G418 or PMG-LEU) never turned red, demonstrating the high repressive activity of Paf1. This explains unsatisfactory results of previous attempts to induce the formation of stable heterochromatin *in trans* by expressing synthetic siRNAs.



Extended Data Figure 8 | Effect of Paf1C mutations on global gene expression and silencing. **a**, The effect of the Paf1C mutations on genome expression was assessed by hybridizing total RNA to whole-genome tiling arrays. The parental wild-type strain, all Paf1C point mutations discovered in the screen, and full deletions of the *paf1+* and *leo1+* genes were included in the analysis. To compare the genome-wide expression profiles of the mutants with the wild-type strain, a principal component analysis (PCA) was performed on the data obtained for two biological replicates of each strain. Principal component (PC) 1 and 2 explained 41.5% and 16.4% of the variance between samples and were selected for visualization, revealing that *cdc73-G313R* and *paf1Δ* cells are most different from wild-type cells. All the other mutants clustered together in a group of samples that also includes wild type, demonstrating that RNA steady-state levels are only minimally affected in these mutants. Note that *leo1Δ* is more similar to wild type than *paf1Δ*, as well as that *paf1Δ* clusters separately from the Paf1C point mutants. **b**, Pairwise comparisons of gene expression between wild-type and *paf1* mutant strains. **c**, *leo1Δ* cells have no growth defect but are susceptible for *de novo* formation of heterochromatin by siRNAs acting *in trans*. These results suggest that Leo1 might be a bona fide repressor of small-RNA-mediated heterochromatin formation.



Extended Data Figure 9 | Kinetic model for Paf1C-mediated repression of siRNA-directed heterochromatin formation. **a**, Paf1C facilitates rapid transcription and release of the nascent transcript from the DNA template. Because the kinetics of transcription termination and RNA 3′ end processing is faster than RITS binding and CLRC recruitment, stable heterochromatin and long-lasting gene silencing cannot be established. **b**, In Paf1C mutant cells identified in this study, elongation of RNA polymerase II, termination of transcription, and the release of the nascent transcript from the site of transcription is decelerated. This results in an accumulation of RNA polymerases that are associated with nascent transcripts, opening up a window of opportunity for the siRNA-guided RITS complex to base-pair with nascent transcripts and recruit CLRC. Consequently, highly stable and repressive heterochromatin is assembled, which is accompanied by the generation of secondary siRNAs covering the entire locus (not depicted in this scheme). Notably, our results demonstrate that impaired transcription termination but not elongation is sufficient to allow silencing. However, to confer robustness to the repressed state, both transcription termination and release of the RNA transcript from the site of transcription must be impaired concomitantly.



Extended Data Figure 10 | Formation of ectopic heterochromatin. **a**, Differential gene expression compared to differential antisense siRNA expression in *leo1*Δ. Gene expression profiles were obtained with whole-genome tiling arrays and small RNA profiles by deep sequencing. Genes neighbouring the *nmt1+::ade6-hp+*, *trp1+::ade6+* and *ade6-704* loci are marked in colour (see also Supplementary Table 1). **b**, siRNA reads mapping to the *ade6-704* locus in *leo1+* and *leo1*Δ strains. Red, plus strand; blue, minus strand. Intronic *rpl2302* siRNAs in *leo1*Δ cells indicate co-transcriptional double-stranded RNA synthesis by RDRC before splicing. **c**, ChIP experiment showing H3K9me2 enrichments on genes surrounding the *ade6-704* locus in *leo1+* and *leo1*Δ cells. Enrichments were calculated relative to background levels obtained in *clr4*Δ cells and normalized to *adh1+*. Error bars, s.d.; mean of $n = 2$ independent biological replicates.

CAREERS

STABILITY National Cancer Institute offers grants for staff scientists **p.255**

FUNDING Grant applications take time and effort, with no guarantee of success **p.255**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

PW ILLUSTRATION/GETTY

DATA SCIENCE

Industry allure

PhD holders with quantitative skills are landing posts at technology companies.

BY MONYA BAKER

Eli Bressert planned to spend his academic career in search of forming stars. He had completed a PhD in astronomy at the University of Exeter, UK, and had won a prestigious postdoctoral fellowship to study radio astronomy near Sydney, Australia. Citations of his papers and invitations for collaborations and conference talks were on the rise. He had no reason to want to work outside astronomy.

But a year into his studies in 2012, the grim reality of the academic job market began to make him nervous. "I sat down and calculated my odds," he recalls. "What was the chance of getting in at a good research institution in a place where my family would be happy?" He had already moved himself, his wife and their year-old son some 16,000 kilometres to Australia for his postdoc, and more transglobal moves for low pay and little stability did not appeal. Still, his research was going well, and he decided to carry on.

That same year, he and a colleague published a handbook on scientific programming, and he was recruited as an academic adviser to a start-up company that was creating software to help collaborators to co-author papers. Bressert loved the energy of the start-up and when he heard of a fellowship that groomed scientists for technology jobs in Silicon Valley, he applied — and was accepted.

He and his family moved again, this time 12,000 kilometres to Palo Alto, California. Today, he is head of data labs at Stitch Fix, a company in San Francisco, California, that creates predictive algorithms that help clients to choose clothes. He says that he loves his work evaluating computational methods in part because it offers more intellectual freedom and creativity than he had experienced in academia.

Bressert is hardly an anomaly — his company employs 20 PhD holders from disciplines as varied as astronomy, neuroscience and electrical engineering. Their biggest asset is rigorous ▶

thinking, says Eric Colson, Bressert's manager. PhD training means learning to formulate questions, test hypotheses and assess whether a solution is reliable. When it comes to modelling data, these qualities make PhD holders more sceptical than most, says Colson. "If it was perfect on the first try, a PhD's first response will be that it is too good to be true. PhDs have this patience and way of framing problems that MBAs don't have." Stitch Fix's PhD holders are just a few of the many young scientists, mainly in the United States, who have left the academic quagmire for jobs in industrial data science.

MAKE THE LEAP

Mathematicians and computer scientists are well-represented in the data-science field, but computing savvy and communication skills matter more than scientific speciality. Early-career researchers hoping to make the transition need to show that they can extract patterns from messy data and place those patterns in the context of commercial goals.

"It's important to remember that industry doesn't value insights. They value analyses that are actionable," says Michael Li, who is co-founder of The Data Incubator, a training course based in New York and Washington DC that prepares graduate students for jobs in data science. And academics skewer their chances by not knowing the ins and outs of industry, says Jake Klamka, who founded a similar training programme, Insight Data Science in Palo Alto. Otherwise qualified candidates can be dismissed as clueless for using the wrong word, such as the academic term 'study' instead of the industry argot 'experiment' or 'A/B test'.

Klamka found it hard to break into industry. He quit his PhD programme in particle physics at the University of Toronto, Canada, in 2010 and began developing tech tools in his kitchen. But although he had the expertise, he lacked knowledge of the industry. "I was 99.5% there in terms of skills," he says. "What I needed was guidance and mentorship." After a year of frustration, he headed to Silicon Valley, where he met software engineers and entrepreneurs who put him on the right track. And thanks in part to backing from the start-up incubator Y Combinator, based in Mountain View, California, he was able to launch his own company, Noteleaf.

Klamka knew that many of his friends in the physics community were interested in moving into industrial data science but were struggling, like he had, to break into industry. At the same time, his tech-community friends were complaining that they had open positions but no one smart enough to fill them. So Klamka founded Insight Data Science to provide PhD holders with the training they need for a career in industrial data science. So far, everyone who has completed the 7-week programme has received job offers (see 'Learn the ropes').


Trainees attend a Science to Data Science workshop in London.

PIVIGO ACADEMY

JOB DESCRIPTIONS

Data-scientist jobs vary widely. Some require mainly tedious 'data munging', cleaning data and filling in gaps to make data sets suitable for relatively simple analysis. Some data scientists work as consultants on data applications; others craft new models and methodologies. Large firms such as LinkedIn, Google and Facebook, with their huge user bases and data sets, tend to support the most sophisticated data modelling.

Would-be data scientists should think broadly about their interests and where they can do what interests them, says Glenn Wong, who has a PhD in physics and is now vice-president at Recorded Future in Somerville, Massachusetts, which organizes web data to help clients deflect cyberattacks. "I don't mean 'how this snippet of DNA interacts with that snippet of DNA'," clarifies Wong, "but 'I like solving problems of a complex two-dimensional nature'. Or 'I like being surrounded by people who have wacky ideas and don't care about hierarchy'."

Amy Heineike took a leave of absence from her PhD programme in computational social science to join a tech start-up based in San Francisco, California, that helps to advise and evaluate early-stage entrepreneurs. "The reason I was doing a PhD was to solve interesting problems, but we were already doing that," she says of her work at the firm. Several years out of academia, and now with stints at other start-ups under her belt, Heineike thinks that she has better opportunities to build ideas and implement them in industry because companies actually connect

LEARN THE ROPES

Find the data-science course to suit you

Many who plan to move into industry use their time at research institutions to burnish their skills and explore their options. Eli Bressert, head of data labs at Stitch Fix in San Francisco, California, recommends learning industry-favoured programming tools such as Python and R. For those who need to boost software skills, programmes such as Data Carpentry and Software Carpentry bring two-day courses to campuses across the world.

Glenn Wong, vice-president of the cybersecurity company Recorded Future in Somerville, Massachusetts, took Harvard Business School seminars when he was a physics PhD student at Harvard University in Cambridge, Massachusetts. The seminars later helped him to cruise through interviews at management-consulting companies.

As a postdoctoral student in synthetic biology at the Massachusetts Institute of Technology in Cambridge, Joy Tharathorn Rimchala, now a data scientist at financial-software company Intuit in Mountain View, California, was uncertain about leaving her academic career until she began auditing a computer-science course. "This is when I decided that data science is cool; at least as cool as my PhD," she says.

Rimchala and Bressert both moved into industry through a programme offered by Insight Data Science in Palo Alto, California. (Last year, a parallel programme opened in New York City and one in Boston will launch in July.) Course attendees work in teams to develop data-driven web applications, and also meet with data scientists at tech firms. The course is free: its costs are met by tech companies, which pay to hire fellows.

A similar initiative is Science to Data Science in London, which offers a 5-week workshop for about 85 students, who pay a £360 (US$540) fee that covers accommodation. After a week and a half of course work, small teams work with mentors from local companies to build practical tools with the companies' own data. Most of last year's fellows returned to their laboratories after finishing the inaugural programme in September, but 75% now have data-science jobs in industry, says co-founder Kim Nilsson, who has a PhD in astrophysics.

Another option is the free, 7-week Data Incubator course, which is based in New York and Washington DC and opens in San Francisco in summer 2015. Finally, the 12-week NYC Data Science Academy programme in New York City, which launched this year, costs $16,000 including coursework on tools such as R, Hadoop and Python. All programmes have more applicants than places. **M.B.**

ACADEMIC DATA DRIVE

Universities create data-science hubs

Academic science, not just industry, has a growing need for data scientists. A US$58-million effort launched last year aims to fill this gap by creating data-science hubs at the University of Washington in Seattle, the University of California, Berkeley (UCB), and New York University. The universities, along with the Gordon and Betty Moore Foundation in Palo Alto, California, and the Alfred P. Sloan Foundation in New York City, are co-funding the hubs. Grants from the Moore Foundation will be given to investigators to develop and refine data-use tools.

Karthik Ram, an assistant researcher at UCB's newly created Berkeley Institute for Data Science, is one of the first beneficiaries. His career advancement depends on his contributions of open-source code and efforts to make data more reproducible, rather than on the conventional criteria for tenure-track posts, such as publication and citation records.

Moore Foundation programme manager Chris Mentzel describes Ram and his colleagues as pioneers in a field that is gaining momentum. "We are trying to create homes for these types of researchers," he says. M.B.

with the people who use the products.

But PhD graduates have to be comfortable with abandoning quests for ever-greater accuracy in favour of commercial goals. Once a data model is working, academics might focus on sophisticated tweaks to improve accuracy and account for outliers. "But in industry, you'd be saying, 'How do I build this into the software; how do I make sure that it won't crash?'" says Heineike. "You have to go the distance for what users really want, and that's something you don't necessarily have time for in academia."

Some hiring managers worry that a desire to craft increasingly accurate models can lead academicians into an unproductive morass. John Baker, who founded a consultancy for data-science services called Datakin in Boston, Massachusetts, recalls an astrophysicist nicknamed 'Dark Matter' by his colleagues because his zeal for perfecting data models meant that he never completed his projects.

David Freeman, head of security data science at the networking firm LinkedIn in Mountain View, says that it is possible to weed out those with such tendencies during interviews. When asked to describe their accomplishments, the most-promising candidates focus more on codes they have implemented than papers they have published. Portfolios developed independently or at boot camps are another good sign of an industry fit, says Baker. "You can tell who is really academic and who really has potential by their projects."

Will Cukierski got noticed this way. He earned his PhD at Rutgers University in New Brunswick, New Jersey, where he taught computers to recognize telltale pathologies in cancerous tissues. But at night, he worked on a challenge from streaming-media provider Netflix: a US$1-million prize to anyone who could best its own movie-recommendation algorithms. He didn't win, but he caught the bug and started to spend his free time on similar contests hosted by the data-science company Kaggle, based in San Francisco. In 2012, company executives contacted him — they had noticed his entries and thought that he could earn a spot on their team. He started there as a data scientist a week after he defended his PhD.

For many PhD holders, the key to success is to find a company whose product or service fascinates them, says Sebastian Gutierrez, author of *Data Scientists at Work*. "You need someone who is excited enough about the business that they actually care that they need to meet quarterly budgets and goals."

Posts for data scientists are starting to emerge in academia (see 'Academic data drive'), but many find the industry environment more appealing. "In industry I can use 20% of the time to achieve 80% of the goal, instead of vice versa," says Shani Offen, formerly a research professor in neuroscience at New York University and now a data scientist at the question-answering site About.com, based in New York. Tommy Guy, a data scientist at the tech giant Microsoft in Bellevue, Washington, likes being rewarded for getting the right answer, no matter what it is. For instance, he can use data analysis to conclude that a proposed new feature would be unpopular with users and argue to dump it, saving the company a considerable sum and earning accolades. Conversely, he says, academia rarely rewards negative results.

Freeman likes the pace at LinkedIn. He recalls doing cutting-edge research in his postdoctoral work at Stanford University in California. "But the thing I was working on would not be seen in actual use for 20 years, if ever. I was looking for something with more immediate impact." And there's nothing like constant deadlines to focus the mind. ■

Monya Baker *writes and edits for* Nature Careers.

GRANTS

Charity for research

Three US philanthropies have pledged a total of US$148 million to support early-career researchers. The Howard Hughes Medical Institute (HHMI), Bill & Melinda Gates Foundation and Simons Foundation will collectively award up to 70 five-year grants of $100,000–400,000 a year, adjusting the amount to account for recipients' other funding. To qualify, applicants must have received a competitive grant within the past 2 years, have 4–10 years' experience in a tenure-track or equivalent position and work at an eligible institution. The Faculty Scholars Program should counter a push towards 'safer', but often less-creative proposals, says HHMI chief scientific officer Erin O'Shea. Applications are due on 28 July.

CAREER PATHS

Staff scientists backed

Non-faculty researchers in university labs and core facilities will soon be able to apply for grants from the US National Cancer Institute in Bethesda, Maryland. The 'research specialist' award addresses the unsustainable growth in the number of postdoctoral positions and reflects an effort to both develop new career paths and provide stability for biomedical labs, says Dinah Singer, director of the institute's division of cancer biology. Lab-research scientists, facility managers and data scientists can work with a sponsoring principal investigator to apply for renewable grants that will cover specialists' salaries and travel. The pilot programme would award 50–60 five-year grants, totalling US$5 million in the first year.

FUNDING

Success takes repetition

A survey of 113 astronomers and 82 psychologists who applied for US federal funding between January 2009 and November 2012 found that applications took on average 116 hours to prepare for principal investigators and 55 hours for co-investigators. More submissions increased the chances of receiving funding, but time spent writing a proposal had little correlation with success. The authors of the survey (T. von Hippel and C. von Hippel. *PLoS ONE* **10,** e0118494; 2015) recommend that investigators avoid programmes with low funding rates unless they are prepared to write two or more proposals a year or have an exceptional application.

ACADEMIC DATA DRIVE

Universities create data-science hubs

Academic science, not just industry, has a growing need for data scientists. A US$58-million effort launched last year aims to fill this gap by creating data-science hubs at the University of Washington in Seattle, the University of California, Berkeley (UCB), and New York University. The universities, along with the Gordon and Betty Moore Foundation in Palo Alto, California, and the Alfred P. Sloan Foundation in New York City, are co-funding the hubs. Grants from the Moore Foundation will be given to investigators to develop and refine data-use tools.

Karthik Ram, an assistant researcher at UCB's newly created Berkeley Institute for Data Science, is one of the first beneficiaries. His career advancement depends on his contributions of open-source code and efforts to make data more reproducible, rather than on the conventional criteria for tenure-track posts, such as publication and citation records.

Moore Foundation programme manager Chris Mentzel describes Ram and his colleagues as pioneers in a field that is gaining momentum. "We are trying to create homes for these types of researchers," he says. M.B.

with the people who use the products.

But PhD graduates have to be comfortable with abandoning quests for ever-greater accuracy in favour of commercial goals. Once a data model is working, academics might focus on sophisticated tweaks to improve accuracy and account for outliers. "But in industry, you'd be saying, 'How do I build this into the software; how do I make sure that it won't crash?'" says Heineike. "You have to go the distance for what users really want, and that's something you don't necessarily have time for in academia."

Some hiring managers worry that a desire to craft increasingly accurate models can lead academicians into an unproductive morass. John Baker, who founded a consultancy for data-science services called Datakin in Boston, Massachusetts, recalls an astrophysicist nicknamed 'Dark Matter' by his colleagues because his zeal for perfecting data models meant that he never completed his projects.

David Freeman, head of security data science at the networking firm LinkedIn in Mountain View, says that it is possible to weed out those with such tendencies during interviews. When asked to describe their accomplishments, the most-promising candidates focus more on codes they have implemented than papers they have published. Portfolios developed independently or at boot camps are another good sign of an industry fit, says Baker. "You can tell who is really academic and who really has potential by their projects."

Will Cukierski got noticed this way. He earned his PhD at Rutgers University in New Brunswick, New Jersey, where he taught computers to recognize telltale pathologies in cancerous tissues. But at night, he worked on a challenge from streaming-media provider Netflix: a US$1-million prize to anyone who could best its own movie-recommendation algorithms. He didn't win, but he caught the bug and started to spend his free time on similar contests hosted by the data-science company Kaggle, based in San Francisco. In 2012, company executives contacted him — they had noticed his entries and thought that he could earn a spot on their team. He started there as a data scientist a week after he defended his PhD.

For many PhD holders, the key to success is to find a company whose product or service fascinates them, says Sebastian Gutierrez, author of *Data Scientists at Work*. "You need someone who is excited enough about the business that they actually care that they need to meet quarterly budgets and goals."

Posts for data scientists are starting to emerge in academia (see 'Academic data drive'), but many find the industry environment more appealing. "In industry I can use 20% of the time to achieve 80% of the goal, instead of vice versa," says Shani Offen, formerly a research professor in neuroscience at New York University and now a data scientist at the question-answering site About.com, based in New York. Tommy Guy, a data scientist at the tech giant Microsoft in Bellevue, Washington, likes being rewarded for getting the right answer, no matter what it is. For instance, he can use data analysis to conclude that a proposed new feature would be unpopular with users and argue to dump it, saving the company a considerable sum and earning accolades. Conversely, he says, academia rarely rewards negative results.

Freeman likes the pace at LinkedIn. He recalls doing cutting-edge research in his postdoctoral work at Stanford University in California. "But the thing I was working on would not be seen in actual use for 20 years, if ever. I was looking for something with more immediate impact." And there's nothing like constant deadlines to focus the mind. ■

Monya Baker *writes and edits for* Nature Careers.

GRANTS

Charity for research

Three US philanthropies have pledged a total of US$148 million to support early-career researchers. The Howard Hughes Medical Institute (HHMI), Bill & Melinda Gates Foundation and Simons Foundation will collectively award up to 70 five-year grants of $100,000–400,000 a year, adjusting the amount to account for recipients' other funding. To qualify, applicants must have received a competitive grant within the past 2 years, have 4–10 years' experience in a tenure-track or equivalent position and work at an eligible institution. The Faculty Scholars Program should counter a push towards 'safer', but often less-creative proposals, says HHMI chief scientific officer Erin O'Shea. Applications are due on 28 July.

CAREER PATHS

Staff scientists backed

Non-faculty researchers in university labs and core facilities will soon be able to apply for grants from the US National Cancer Institute in Bethesda, Maryland. The 'research specialist' award addresses the unsustainable growth in the number of postdoctoral positions and reflects an effort to both develop new career paths and provide stability for biomedical labs, says Dinah Singer, director of the institute's division of cancer biology. Lab-research scientists, facility managers and data scientists can work with a sponsoring principal investigator to apply for renewable grants that will cover specialists' salaries and travel. The pilot programme would award 50–60 five-year grants, totalling US$5 million in the first year.

FUNDING

Success takes repetition

A survey of 113 astronomers and 82 psychologists who applied for US federal funding between January 2009 and November 2012 found that applications took on average 116 hours to prepare for principal investigators and 55 hours for co-investigators. More submissions increased the chances of receiving funding, but time spent writing a proposal had little correlation with success. The authors of the survey (T. von Hippel and C. von Hippel. *PLoS ONE* **10,** e0118494; 2015) recommend that investigators avoid programmes with low funding rates unless they are prepared to write two or more proposals a year or have an exceptional application.

BREAD OF LIFE

Home truths.

BY BETH CATO


ILLUSTRATION BY JACEY

It was dangerous — traitorous — for people to speak aloud of their memories of Earth. It meant they risked giving the Dendul exactly what they wanted. And yet, people often couldn't help but dismiss that peril when they entered Sonya's Earth Bread Shop.

Sonya looked up as a man entered her business. He was the typical sort. In his forties or fifties, his hair silver; not that different from her, gender aside. Both of them old enough to remember Earth in its glory. He paused at the door and breathed in. Grief flashed across his face.

"My god," he whispered.

"You must be Franklin with the order for rye. Your ship docked right on time."

Humanity had scattered across a dozen systems, yet Sonya didn't have to advertise her wares. Word managed to spread among any human crew members on the freighters and shuttles that passed through Kaji Station. Most ordered ahead to get what they wanted, but she always kept favourites on display. Bread rarely went to waste.

A thin visor wrapped around Franklin's eyes like a clear halo and did nothing to hide his dazed expression. "This place — how do you even get the ingredients?" He had a trim body like most deep spacers, as they relied on strict calorie packets when in transit.

"It's expensive. I don't make much profit." That was the truth. "I source my wheat and everything else through legitimate traders who specialize in human palatable foods. I beamed you that info when you ordered."

Sonya pulled out a bundle wrapped in white parchment. The paper was folded just so, as if she had swaddled a newborn baby.

"I read it. Maybe it seems too good to believe." He stared at the goods preserved beneath the counter's dome. "You have sourdough available in slices? And challah?" He leaned against the glass as if suddenly boneless.

Sonya stared at him, waiting.

"My grandmother. She used to make challah for holidays." He said it in the lowest discernible whisper. "This big braided round. She insisted it had to be done by hand, that it tasted best. The smell, when that loaf would come out of the oven." He inhaled, his breath rattling with checked emotion. "It even smells right here. I had forgotten. It's been so long since …"

The Dendul had obliterated Earth.

To them, humans radiated potent emotions in a way unlike any other species. Spoken memories, in particular, exuded deep flavours that the Dendul absorbed to ascend into a state of blissful intoxication. When they scorched Earth, their intent wasn't to slaughter the majority of humanity. No, they ripened memories for harvest.

"Most every culture on Earth had some kind of bread," said Sonya. "I make it all. I hear it all. I think almost everyone had a grandmother who baked. I even have people who come to me for those old commercial sliced breads."

He laughed, high and giddy. "The stuff in every kid's lunch box, with peanut butter and jelly, or that god-awful bologna. Even that sounds good these days. This challah in here? How much —" He blinked as she beamed the new total to his visor. The non-vocal response was habit for Sonya; the Kaji regarded the mention of money as crude. "I'll take the rest of that loaf. There's a woman on my crew … this will mean a lot to her."

Sonya wrapped it up. "Thanks for beaming your payment promptly."

He nodded as she passed the bagged bread to him. "Yeah. Yeah. You do this … why? Doesn't it drive you crazy to remember what we lost, every day?"

"It would drive me crazier not to."

He backed away, his gaze distant, and said nothing as he departed.

Sonya looked up at the obscured sensors that recorded their conversation at much deeper levels than mere vocalization. A human with synaesthesia might hear music and see colours; for the Dendul, Franklin's brief, emotional story would evoke catatonic ecstasy.

The Dendul had tried torturing humans to force out memories, but physical pain tainted the results. Made them bitter. Alcohol also changed the flavour in a distasteful way. The words had to be provoked in an unimpaired, natural way.

Many Earth foods proved to be good bait, but nothing was as powerful and universal as bread.

Sonya reached into the flash freezer for another loaf of challah for the dome, and to ready her next orders for the day. Portuguese sweet bread. Pitta. Ciabatta. She breathed in the redolent, yeasty smell.

This memory trap was for her own selfish needs, too. To draw in the few humans who came to this far station, to hear them whisper history as a conspiracy. She drank it in. Savoured it. As if she still ran the counter at her father's bakery in Denver. As if she wasn't on a revolving rod in deep space, a speck in a diaspora of a million remaining human souls.

The door chimed and a white-haired woman entered. "There really is a bakery on this station! The smell here. It reminds me …" She drew silent, her expression pained.

Sonya said nothing. She waited. The stories would come. ■

Beth Cato *resides in Arizona. She is the author of the* Clockwork Dagger *steampunk fantasy series from Harper Voyager. Her website is BethCato.com.*

➲ NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm